



INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Stark:

Zu Vorhersage und Vorhersagewert fuer Ueberlebenszeitmodelle

Sonderforschungsbereich 386, Paper 84 (1997)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Zu Vorhersage und Vorhersagewert für Überlebenszeitmodelle

MONIKA STARK¹

*Institut für Medizinische Statistik und Epidemiologie
Technische Universität München*

Zusammenfassung

Neben der Modellanpassung und der Signifikanz der Kovariablen besteht von Anwenderseite verstärktes Interesse an Aussagen über den Vorhersagewert eines Modells für neue Patienten und den bisher gewonnenen Erklärungsgrad der Kovariablen für die Zielgröße. Die für Überlebenszeitmodelle aus verschiedenen Ansätzen stammenden Definitionen und Schätzer für entsprechende Maßzahlen werden auf eine allgemeine Form zurückgeführt beziehungsweise aus dieser hergeleitet und gegenübergestellt. Unter Verwendung von Martingal-Residuen wird ein neuer Ansatz angegeben, der zensierte Beobachtungen mit einbezieht, ohne für deren Verluste zu extrapolieren.

¹email: monika.stark@imse.med.tu-muenchen.de

1 Einleitung

Für Überlebenszeitmodelle sind die für andere Modelle bekannten Methoden zur Schätzung der Vorhersage und des Vorhersagewertes nicht direkt anwendbar. Dies gilt speziell für nonparametrische Ansätze. Meist wird eine Metrik verwendet, die den Abstand (Verlust, Fehler) zwischen vorherzusagender Zielgröße und beobachteten Werten dieser Größe mißt. Wegen der Berücksichtigung unvollständiger Beobachtungen ist eine solche Metrik jedoch weder bezüglich ihrer Schätzung noch ihrer Interpretation übertragbar. Zielgröße "Ausfallzeit" und beobachtete Größe "Beobachtungszeit" stimmen nicht überein. Es müssen Modifikationen und andere Ansätze verwendet werden, die die Struktur dieser Daten in geeigneter Weise berücksichtigen.

Ein einfaches Anwendungsbeispiel ist ein Datensatz von 179 Magenkarzinompatienten nach vollständiger Resektion des Tumors einer prospektiven Studie der Chirurgischen Klinik der Technischen Universität München. Um die Vorhersage bezüglich Rezidivrisiko, Überlebenszeit oder beschreibenden Parametern zu verbessern, lassen sich die Patienten in zwei Risikogruppen nach Risikofaktor "Prozentsatz befallener Lymphknoten über bzw. unter 13% " einteilen (Abb.1). Für die Schätzung von Vorhersage und Vorhersagewert sind die Zensierungen geeignet zu berücksichtigen und interpretierbare Maßzahlen zu verwenden.

2 Vorhersage und Verlustfunktionen

Eine Vorhersage kann gemacht werden für Beobachtungen aus dem vorliegenden Datensatz oder für zukünftige Beobachtungen aus der Grundgesamtheit.

Vorhergesagt wird die Zielgröße, oft auch Parameter der empirischen Verteilung oder unter einer Verteilungsannahme.

Im parametrischen Kovariablenmodell werden Regressionsparameter für den Einfluß der Kovariablen auf die Zielgröße oder auf einen Parameter, der die Zielgröße (mit)bestimmt, vorhergesagt.

Vorhersagefehler

Allgemein kann die beste Vorhersage bezüglich der Überlebenszeit T bestimmt

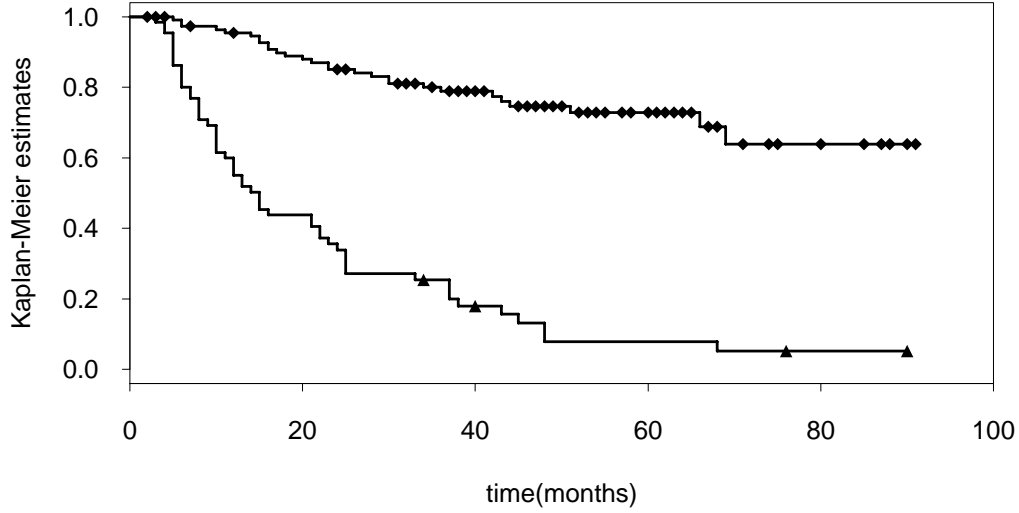


Abbildung 1: Hoch- und Niedrig-Risikogruppe Magenkarzinompatienten

werden, indem der erwartete Fehler oder auch Verlust der Vorhersage, die mit Hilfe des Modells gemacht werden kann, minimiert wird (VanHouwelingen&LeCessie90, Korn&Simon90)

$$\min_{\tilde{t}} E(L(T, \tilde{t}), S(\cdot)) = \min_{\tilde{t}} \int_0^{\infty} L(t, \tilde{t}) d(1 - S(t)) \Rightarrow \tilde{t}_{opt}$$

Die Verlustfunktion L als Funktion der Zielgröße T und des Prädiktors \tilde{t} bestimmt Charakter und Struktur des Prädiktors, $S(\cdot)$ dessen Verteilung und Wert.

Werden erklärende Kovariablen X bei der Vorhersage verwendet, lautet der zu minimierende erwartete Verlust, gegeben x

$$\min_{\tilde{t}} E(L(T, \tilde{t}), S(\cdot|x)) = \min_{\tilde{t}} \int_0^{\infty} L(t, \tilde{t}) d(1 - S(t|x)) \Rightarrow \tilde{t}_{opt}|x$$

Je nach Anwendung können unterschiedliche Verlustfunktionen zu einem anwendungsspezifisch optimalen Prädiktor führen. Es führt etwa die übliche quadratische Form der Verlustfunktion zur Definition des *mean squared error* als

erwartetem Fehler

$$MSE := E((T - \tilde{t})^2, S(\cdot))$$

und dieser allgemein zum optimalen Prädiktor $\tilde{t}_{opt} = \int_0^\infty td(1 - S(t)) = E(T)$, dem Erwartungswert von T .

Der minimale, erwartete Fehler lautet

$$E(L(T, \tilde{t}_{opt}), S(\cdot))$$

und ergibt für den MSE die Varianz von T , $E((T - \tilde{t}_{opt})^2) = E((T - E(T))^2) =: Var(T)$.

Da die Überlebenszeit meist schief verteilt ist, gilt der Erwartungswert oft nicht als geeigneter Prädiktor. Eine naheliegende, einfache Alternative liefert der absolute Fehler als Verlustfunktion $L(t, \tilde{t}) = |t - \tilde{t}|$. Er führt zu $\tilde{t}_{opt} = S^{-1}(0.5)$, dem Median von T , mit dem erwarteten Fehler $E(|T - S^{-1}(0.5)|)$.

Weiterhin wurde vorgeschlagen, für den Fehler beobachtete und vorhergesagte Zeit zu logarithmieren, um eine Abweichung der Vorhersage bei fortgeschrittener Überlebenszeit schwächer zu bewerten als die gleiche Abweichung bei niedriger Überlebenszeit (Transformationen innerhalb der Verlustfunktion bezüglich der Lage des Fehlers)

$$L(t, \tilde{t}) = (\ln t - \ln \tilde{t})^2$$

Es folgt $\exp(E(\ln T))$ als Prädiktor und $Var(\ln T)$ als erwarteter Fehler.

Es kann ebenso eine Zensierung des Verlustes ab einer gewissen Beobachtungsdauer t_0 angezeigt sein, wenn etwa für $t \geq t_0$ eine Vorhersage mit $\tilde{t} \geq t_0$ als vollständig korrekt gelten soll. Zum besten Prädiktor und dessen Fehler führt eine Substitution von t durch $t^* := \min(t, t_0)$ (Korn&Simon90)

$$L_{t_0}(t, \tilde{t}) := L(t^*, \tilde{t}_{t_0})$$

Es ergibt sich bei quadratischer Verlustfunktion als Fehler

$$\int_0^\infty L_{t_0}(t, \tilde{t})d(1 - S(t)) = \int_0^\infty L(t^*, \tilde{t}_{t_0})d(1 - S(t^*))$$

und für den Prädiktor $\tilde{t}_{t_0_{opt}} = E(T^*) = E(\min(T, t_0))$. Als Fehler folgt $Var(T^*) = Var(\min(T, t_0))$.

Bei Verwendung des absoluten Fehlers ergibt sich $\tilde{t}_{t_{0_{opt}}} = \min(\text{Med}(T), t_0)$ und für den erwarteten Fehler $E(|\min(T, t_0) - \tilde{t}_{t_{0_{opt}}}|)$.

Grundsätzlich gilt $\tilde{t}_{t_{0_{opt}}} \leq \tilde{t}_{opt}$ und $\tilde{t}_{t_{0_{opt}}} \leq t_0$, wobei $\tilde{t}_{t_{0_{opt}}} = t_0$ als $\tilde{t}_{t_{0_{opt}}} \geq t_0$ zu interpretieren ist (Zensierung der Verlustfunktion bezüglich der Lage des Fehlers).

Andere Kriterien können ein Abfall in der Bewertung des Fehlers mit zunehmender Größe des Fehlers sein. Bei ohnehin großen Abweichungen zwischen Beobachtung und Vorhersage sind die Verluste nicht so stark zu unterscheiden wie bei geringen Abweichungen (Abfall der Verlustfunktion bezüglich der Größe des Fehlers).

Außerdem kann eine Überschätzung der Überlebenszeit gewichtiger sein als eine Unterschätzung der Überlebenszeit (Asymmetrie der Verlustfunktion bezüglich der Richtung des Fehlers). Dies kann ebenso umgekehrt gelten. Eine flexiblere Formulierung einer Verlustfunktion lautet (Henderson95)

$$L^H(t, \tilde{t}) := \begin{cases} \frac{aL(t, \tilde{t})}{b_1 + L(t, \tilde{t})} & : \tilde{t} < t \\ \frac{(1-a)L(t, \tilde{t})}{b_2 + L(t, \tilde{t})} & : \tilde{t} \geq t \end{cases}$$

Der Parameter a repräsentiert hierbei die Asymmetrie der Bewertung bezüglich der Richtung des Fehlers mit $0 \leq a \leq 1$, wobei $a < 0.5$ eine Überschätzung der Überlebenszeit stärker gewichtet als eine Unterschätzung. b_1 und b_2 bestimmen die Abhängigkeit von der Größe des Fehlers für positive beziehungsweise negative Abweichungen. $L(t, \tilde{t})$ gibt eine quadratische Verlustfunktion an wie zum Beispiel

$$L(t, \tilde{t}) := (\sqrt{t} - \sqrt{\tilde{t}})^2$$

, die eine Wurzeltransformation der Überlebenszeit für eine niedrigere Gewichtung von später auftretenden und gleichzeitig von größeren Abweichungen enthält.

Eine geringe Asymmetrie bei der Gewichtung positiver und negativer Abweichungen der Vorhersage verursacht erhebliche Abweichungen bei der Wahl der besten Vorhersage. Bei symmetrischer Gewichtung der Abweichungen mit $a = 0.5$ ist für die vollständige Definition der Verlustfunktion nur ein nicht normierter Parameter $b_1 = b_2 = b$ für den jeweiligen Datensatz gesondert zu wählen (für $b = L(t, \tilde{t})$ folgt $L^H(t, \tilde{t}) = \frac{1}{4}$).

Binäre Verlustfunktionen

Unter Umständen ist nicht die Ausfallzeit, sondern der Status "Ausfall noch nicht eingetreten" zu einem bestimmten Zeitpunkt t_b von Interesse.

Mit Hilfe binärer Verlustfunktionen kann die Überlebenswahrscheinlichkeit zu einem bestimmten Zeitpunkt t_b , also $S(t_b)$, vorhergesagt werden. Diese wird in Beziehung gesetzt zum tatsächlichen Status, der zum Zeitpunkt t_b erreicht wurde, wenn die realisierte Ausfallzeit t ist.

$$I_{\{t > t_b\}} := \begin{cases} 1 & : t > t_b \\ 0 & : t \leq t_b \end{cases}$$

Dieser Status kann als Überlebensfunktion zum Zeitpunkt t_b , gültig für die Realisation t , betrachtet werden, $S_t(t_b)$.

Im Gegensatz zur Überlebenszeit bildet hier der Erwartungswert eine sinnvolle Vorhersage. Die quadratische binäre Verlustfunktion (Korn&Simon90) für den Zeitpunkt t_b lautet in einer anderen Formulierung

$$L(t, S(\tilde{t}_b)) := (I_{\{t > t_b\}} - S(\tilde{t}_b))^2$$

Der optimale Prädiktor $S(\tilde{t}_b)_{opt} = \tilde{S}_{opt}(t_b)$ ist $E(I_{\{T > t_b\}}) = E(S_T(t_b)) = S(t_b)$. Ist t_b der Median von T , ergibt sich für die optimale Vorhersage des Status also 0.5. Der erwartete Fehler lautet

$$\begin{aligned} \int_0^\infty L(t, S(t_b)) dF(t) &= \int_0^\infty (I_{\{t > t_b\}} - S(t_b))^2 dF(t) \\ &= \int_0^\infty (I_{\{t > t_b\}} - S(t_b))^2 dF(I_{\{t > t_b\}}) \\ &= \int_0^\infty (S_t(t_b) - S(t_b))^2 dF(S_t(t_b)) \\ &= \text{Var}(S_T(t_b)) \\ &= S(t_b)(1 - S(t_b)) \end{aligned}$$

Interessieren die Überlebenswahrscheinlichkeiten im gesamten Zeitverlauf, kann die binäre Verlustfunktion über alle möglichen Zeitpunkte t_b addiert beziehungs-

weise integriert werden

$$\int_0^{\infty} L(t, \tilde{S}(t_b)) dt_b = \int_0^{\infty} (I_{\{t > t_b\}} - \tilde{S}(t_b))^2 dt_b$$

Insgesamt wird dann statt einer Verlustfunktion, mit der der Zeitpunkt des Übergangs in den Status "Ausfall" vorhergesagt wird, $L(t, \tilde{t})$, oder nur der erreichte Status zu einem bestimmten Zeitpunkt, $L(t, \tilde{S}(t_b)) = L(S_t(t_b), \tilde{S}(t_b))$, nun eine Verlustfunktion verwendet, mit der der Status im Verlauf der Zeit, also die gesamte Überlebensfunktion vorhergesagt wird, " $L(t, \tilde{S}(\cdot)) = L(S_t(\cdot), \tilde{S}(\cdot))$ " mit $\tilde{S}_{opt}(\cdot) = S(\cdot)$.

Der erwartete integrierte quadratische binäre Fehler zur Bestimmung des Fehlers bei optimalem Prädiktor für alle t_b lautet dann

$$\begin{aligned} & \int_0^{\infty} \left(\int_0^{\infty} L(t, S(t_b)) dt_b \right) dF(t) \\ &= \int_0^{\infty} \left(\int_0^{\infty} L(t, S(t_b)) dF(t) \right) dt_b \\ &= \int_0^{\infty} S(t_b)(1 - S(t_b)) dt_b \end{aligned}$$

Bei der Berechnung des Fehlers ist eine zusätzliche Gewichtung der Zeitpunkte t_b bei der Integration in Abhängigkeit von der Verteilung von T möglich und aus inhaltlichen oder numerischen Gründen sinnvoll (siehe "Schätzung der erklärten Variation"). Ein Fehler bei der Schätzung der Überlebenswahrscheinlichkeit bei niedrigen Überlebenszeiten könnte gewichtiger sein als bei fortgeschrittenen Überlebenszeiten (Gewichtung der Verlustfunktion bezüglich der Lage des Fehlers). Der Prädiktor ändert sich jedoch nicht.

Parametrische und semiparametrische Verlustfunktionen

Eine andere Art von Verlustfunktionen stellt jeweils eine parametrische Verteilungsklasse dar und ergibt sich aufgrund folgenden Zusammenhangs: Im Falle der quadratischen Verlustfunktion einerseits und der Normalverteilung andererseits ist in der negativen (log-)Likelihood bei $-l(\mu, \sigma^2|t)$ mit $\mu = \tilde{t}$ die quadratische

Verlustfunktion $(t - \hat{t})^2$ enthalten

$$-l(\mu, \sigma^2|t) = \ln(\sigma * \sqrt{2\pi}) + \frac{1}{2\sigma^2}(t - \mu)^2$$

Sind nicht Zeit oder Status vorherzusagen, sondern Parameter einer angenommenen Verteilungsklasse der Zielgröße, kann die negative log-Likelihood $-l(\theta|t)$ als Verlustfunktion für Likelihood-basierte Modelle verwendet werden, die statt einer Metrik zwischen der Zielgröße und der vorherzusagenden Größe den angenommenen Zusammenhang zwischen der Zielgröße und den gesuchten Parametern enthält

$$L(t, \tilde{\theta}) = -l(\tilde{\theta}|t)$$

Die gewählte parametrische Verteilungsklasse bestimmt die Struktur des Prädiktors, hier des Parameters $\tilde{\theta}$. Stellvertretend für den erwarteten Fehler und verallgemeinernd für den *MSE* lautet die erwartete (negative) log-Likelihood (Fraser65)

$$\begin{aligned} -F(\tilde{\theta}) &:= E(-l(\tilde{\theta}|T), S(\cdot)) = \int_0^\infty -l(\tilde{\theta}|t)d(1 - S(t)) \\ &= \int_0^\infty -\ln f(t|\tilde{\theta})f(t)dt \end{aligned}$$

Der optimale Prädiktor ist das theoretische Analogon des *ML*-Schätzers; wird durch die Verlustfunktion $l(\tilde{\theta}|t)$ nach den Parametern $\tilde{\theta}$ für eine vorgegebene parametrische Form gesucht, und ist die wahre Verteilung von T bereits von der gleichen parametrischen Form, $f(t) = f(t|\theta)$, dann gilt $\tilde{\theta}_{opt} = \theta$.

Es ergibt sich statt wie bei quadratischer Verlustfunktion $\tilde{t}_{opt} = E(T, f(\cdot))$, jetzt bei negativer log-Likelihood der Normalverteilung $\tilde{\mu}_{opt} = E(T, f(\cdot))$ und $\tilde{\sigma}_{opt}^2 = Var(T, f(\cdot))$.

Der erwartete minimale Fehler bei optimaler Vorhersage $E(T, f(\cdot))$ ergibt allerdings nur bei bekanntem σ^2 die Varianz $Var(T, f(\cdot))$ – in einer linearen Transformation – wie bei der quadratischen Verlustfunktion

$$E(-l(\tilde{\mu}_{opt}|T, \sigma^2), f(\cdot)) = E\left(\ln(\sigma * \sqrt{2\pi}) + \frac{1}{2\sigma^2}(T - E(T, f(\cdot)))^2, f(\cdot)\right)$$

$$\begin{aligned}
&= \ln(\sigma * \sqrt{2\pi}) * E[(T - E(T, f(\cdot)))^2, f(\cdot)] \\
&= \ln(\sigma * \sqrt{2\pi}) * Var(T, f(\cdot))
\end{aligned}$$

Ist σ^2 nuisance-Parameter und $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}^2)$ anzugeben, folgt als Fehler

$$\begin{aligned}
E(-l(\tilde{\mu}_{opt}, \tilde{\sigma}_{opt}^2|T), f(\cdot)) &= E\left(\ln(\sqrt{Var(T, f(\cdot))} * \sqrt{2\pi}) + \frac{1}{2} \frac{(T - E(T, f(\cdot)))^2}{Var(T, f(\cdot))}, f(\cdot)\right) \\
&= \ln(\sqrt{Var(T, f(\cdot))} * \sqrt{2\pi}) + \frac{1}{2}
\end{aligned}$$

Jedoch gilt

$$\exp[2E(-l(\tilde{\mu}_{opt}, \tilde{\sigma}_{opt}^2|T), f(\cdot))] = Var(T, f(\cdot)) * 2\pi * e$$

Bei Annahme der Exponentialverteilung mit

$$-l(\lambda|t) = -(\ln \lambda - \lambda t)$$

folgt für den Parameter $\lambda \tilde{\lambda}_{opt} = \frac{1}{E(T, f(\cdot))}$.

Im Kovariablenmodell ergibt sich für die zu minimierende erwartete negative log-Likelihood, gegeben x

$$\begin{aligned}
-F(\tilde{\theta}|x) := E(-l(\tilde{\theta}|T), S(\cdot|x)) &= \int_0^\infty -l(\tilde{\theta}|t) d(1 - S(t|x)) \\
&= \int_0^\infty -\ln f(t|\tilde{\theta}) f(t|x) dt
\end{aligned}$$

, woraus sich ein $\tilde{\theta}_{opt}|x$ ergibt.

Wird nach einem parametrischen Kovariablenmodell gesucht mit $\theta = (\theta \setminus \beta, \beta)$ und soll β konstant sein bezüglich X , kann dies direkt in die Verlustfunktion mit aufgenommen werden, $-l(\tilde{\theta}|t, x)$, und der Erwartungswert der negativen log-Likelihood über alle x gemeinsam gebildet werden, gegeben die marginale Verteilung $G(x)$ von X (Kent83)

$$\begin{aligned}
-F_c(\tilde{\theta}) := E_x[E(-l(\tilde{\theta}|T, x), S(\cdot|x))] &= E_x\left[\int_0^\infty -l(\tilde{\theta}|t, x) d(1 - S(t|x))\right] \\
&= \int_{-\infty}^\infty \int_0^\infty -\ln f(t|\tilde{\theta}, x) f(t|x) dt dG(x)
\end{aligned}$$

Es ergeben sich unter den für alle Elemente geltenden Parametern $\tilde{\theta}_{opt}$ auch die Effekte der Kovariablen, $\tilde{\beta}_{opt} = \beta$.

Semiparametrische Verlustfunktionen

Bei Vorhersage allein der Effekte und Verwendung des Kovariablen-Modells von Cox könnte die volle allgemeine *PH*-Likelihood (ohne Zensierung) zur Formulierung einer semiparametrischen Verlustfunktion verwendet werden. Es kann dann die erwartete negative log-likelihood formuliert werden, die jedoch wegen der enthaltenen, zeitabhängigen Baseline-Hazardrate $\lambda_0(t)$ nicht in geschlossener Form angegeben werden kann.

Ebenfalls käme hierzu die Partial-Likelihood in Frage. Für diese kann jedoch weder eine Verlustfunktion, noch deren Erwartungswert formuliert werden.

Eine Möglichkeit, um dennoch eine Verlustfunktion und den Erwartungswert des Verlustes angeben zu können, der sich auf die Effekte in einem Cox-Modell beziehen soll, besteht in der Spezialisierung der allgemeinen *PH*-Likelihood beziehungsweise der Partial-Likelihood auf eine Weibullverteilung. Bei geeigneter Transformation der Zeit beziehungsweise Wahl der Baseline-Hazardrate folgt für T ein Weibullverteilungsmodell, das die Kovariableneffekte $\tilde{\beta}$ aus dem Cox-Modell enthält und zusätzlich statt $\lambda_0(t)$ zwei zeitunabhängige Parameter, die das Weibullregressionsmodell vervollständigen, α^* und a^* . So könnte eine erwartete negative log-Likelihood formuliert werden. Weiterhin folgt $\ln T$ einer Extremwertverteilung oder auch einem linearen Regressionsmodell mit standard-extremwertverteiltem Fehlerterm (siehe "Definitionen und Modelle"). Deshalb kann die erwartete Likelihood auch unter Verwendung der Dichte der Extremwertverteilung angegeben und anschließend eine geschlossene Form, abhängig von α^* und a^* , für sie berechnet werden (Kent&O'Quigley88)

$$-F_c(\tilde{\theta}^*) := \int_{-\infty}^{\infty} \int_0^{\infty} -\ln f(t|(\tilde{\beta}, \alpha^*, a^*), x) f(t|x) dt dG(x)$$

3 Schätzung der Vorhersage

Im Allgemeinen ist $S(t|x)$ zur Bestimmung von $\tilde{t}_{opt}|x$ nicht bekannt. Die Verteilung ist mindestens nicht vollständig bekannt, das heißt, das Modell und die

Verteilungsklasse von T werden als bekannt angenommen, die oder der Parameter $\theta|x$ sind jedoch unbekannt. Um in beiden Fällen den besten Prädiktor $\tilde{t}_{best}|x$ für zukünftige Beobachtungen zu finden, ist die geschätzte Verteilung der zukünftigen Beobachtungen zu verwenden.

Ist die Verteilungsklasse von T unbekannt, wird die empirische Verteilung $\bar{S}(t|x)$ verwendet. Bei bekannter Verteilungsklasse ist $\theta|x$ zu schätzen, bei einem parametrischen Kovariablenmodell die Parameter θ , welche die Kovariableneffekte β enthalten. Dabei wird von festen Ausprägungen der Kovariablen ausgegangen, für $dG(x)$ wird die empirische Verteilung von X verwendet, um dann die parametrisch geschätzte Verteilung $\hat{S}(t|x)$ angeben zu können.

Anschließend ist der beste Prädiktor $\tilde{t}_{best}|x$ aufgrund der Verlustfunktion aus der Verteilung oder für allgemeine Verlustfunktionen, wenn keine geschlossene Form des Prädiktors angegeben werden kann, numerisch durch Minimierung des erwarteten Fehlers zu bestimmen.

3.1 Verteilungsfreie Schätzung der Vorhersage

Der gesuchte Prädiktor bei quadratischer Verlustfunktion kann als Erwartungswert der empirischen Verteilung angegeben werden, $\tilde{t}_{best} = E(T, \bar{S}(\cdot)) = \bar{t}$.

Für allgemeine Verlustfunktionen kann der erwartete Fehler verteilungsfrei geschätzt und minimiert werden unter Verwendung der empirischen Verteilungsfunktion $\bar{F}(t)$ beziehungsweise $\bar{S}(t)$,

$$E(L(T, \tilde{t}), \bar{S}(\cdot)) = \sum_{t_i} L(t_i, \tilde{t}) [\bar{S}(t_i^-) - \bar{S}(t_i)] \Rightarrow \tilde{t}_{best}$$

, wobei t^- "unmittelbar vor t " bedeuten soll.

Ohne Zensierungen und für stetige T handelt es sich bei der empirischen Verteilungsfunktion um eine gleichmäßige Stufenfunktion. Deshalb folgt

$$\begin{aligned} E(L(T, \tilde{t}), \bar{S}(\cdot)) &= \sum_{t_i} L(t_i, \tilde{t}) \frac{1}{n} \\ &= \frac{1}{n} \sum_{t_i} L(t_i, \tilde{t}) \end{aligned}$$

Für den MSE etwa ergibt sich dann als Verlust die mittlere Summe der quadratischen Abweichungen (Im Weiteren gelte: $\Sigma = \sum_{i=1}^n$)

$$mse := \frac{1}{n} \sum (t_i - \tilde{t})^2$$

Es folgt ebenso $\tilde{t}_{best} = \bar{t}$.

Unter Verwendung von Kovariablen ergibt sich für den erwarteten Fehler, gegeben x , mit n_x der Anzahl und $\frac{1}{n_x}$ der empirischen Verteilung der Elemente in der Kategorie x

$$E(L(T, \tilde{t}), \bar{S}(\cdot|x)) = \frac{1}{n_x} \sum_{t_i|x_i=x} L(t_i, \tilde{t}) \Rightarrow \tilde{t}_{best}|x$$

Für den quadratischen Fehler folgt $\tilde{t}_{best}|x = \bar{t}|x$.

Bei kategorialen X , das heißt wenn eine geringe Anzahl von möglichen Vorhersagen zu bestimmen ist, kann diese Lösung verwendet werden.

Für stetige X würde sich jedoch für jedes feste x_i $\tilde{t}_{best}|x_i = t_i$ ergeben; eine parametrische Lösung unter Verteilungsannahme oder eine semiparametrische Lösung ist notwendig.

Binäre Verlustfunktionen

Wird die quadratische binäre Verlustfunktion verwendet mit vorherzusagendem Status zu einem bestimmten Zeitpunkt, $S(\tilde{t}_b)$, so gilt $S(\tilde{t}_b)_{best} = \tilde{S}_{best}(t_b) = \bar{S}(t_b)$. Der Prädiktor kann ebenfalls numerisch gefolgert werden, also ohne Verwendung der geschätzten Verteilung \bar{S} . Aus der Minimierung von

$$\frac{1}{n} \sum_{t_i} L(t_i, S(\tilde{t}_b)) = \frac{1}{n} \sum_{t_i} (I_{\{t_i > t_b\}} - \tilde{S}(t_b))^2$$

ergibt sich $\frac{1}{n} \sum_{t_i} I_{\{t_i > t_b\}}$.

Analog folgt $\tilde{S}_{best}(t_b|x)$ aus

$$\frac{1}{n_x} \sum_{t_i|x_i=x} (I_{\{t_i > t_b\}} - \tilde{S}(t_b|x))^2$$

Sind die Überlebenswahrscheinlichkeiten für den gesamten Zeitverlauf vorherzusagen, so sollte dieser auf den Bereich der erhobenen Daten bis t_{max} beschränkt

werden. Die gesamte empirische Verteilung liefert dann den Prädiktor, $\bar{S}(\cdot)$. Wiederum kann die Lösung auch numerisch bestimmt werden, wobei die Verwendung der über t_b von 0 bis t_{max} integrierten binären Verlustfunktion – auch bei Verwendung einer anderen als der Lebesgue-Integration – bei verteilungsfreier Schätzung keine abweichenden Prädiktoren von den für einzelne Zeitpunkte berechneten liefert.

Parametrische und semiparametrische Verlustfunktionen

Bei formal verteilungsfreier Schätzung der Vorhersage im Rahmen einer parametrischen Verlustfunktion folgt für $\tilde{\theta}_{best}$ der ML -Schätzer von θ

$$E_x E(-l(\tilde{\theta}|T, x), \bar{F}(\cdot|x)) = \sum_{t_i, x_i} -\ln f(t_i|\tilde{\theta}, x_i) \frac{1}{n} =: -l \frac{1}{n}(\tilde{\theta})$$

Bei der allgemeinen PH -Likelihood als Verlustfunktion kann für die verteilungsfreier Schätzung der Vorhersage der Effekte die PL verwendet werden. Sie enthält die voneinander abhängigen Komponenten der semiparametrischen negativen log- PL

$$E_x E(-pl(\tilde{\beta}|T, x), \bar{F}(\cdot|x)) \neq -\frac{1}{n} \sum_{i=1}^n \left((x'_i \tilde{\beta}) - \ln \sum_{l \in R(t_i)} \exp(x'_l \tilde{\beta}) \right) =: -\frac{1}{n} pl(\tilde{\beta})$$

Eine zugrunde liegende Verlustfunktion kann jedoch nicht angegeben werden, weshalb unklar ist, welche Größe durch $-\frac{1}{n} pl(\tilde{\beta})$ geschätzt wird.

3.2 Schätzung der Vorhersage bei Zensierungen

Sind aufgrund der Verlustfunktion Größen wie der Median oder Erwartungswert als \tilde{t}_{best} zu berechnen, können diese mit Hilfe der empirischen Verteilung bestimmt werden. Erreicht die Überlebensfunktion 0.5 beziehungsweise 0 in diesen Fällen nicht, kann die entsprechende zensierte Verlustfunktion mit $t_0 = t_{max}$ zu einer Lösung für \tilde{t}_{best} in Form einer Ungleichheitsrelation $\tilde{t}_{best} \geq \tilde{t}_{t_0_{best}}$ führen (wenn nicht bereits eine Zensierung mit $t_0 \leq t_{max}$ in der Verlustfunktion enthalten war). Eine rein spekulative Extrapolation der KM -Überlebensfunktion $\bar{S}(\cdot|x)$ über den Bereich der erhobenen Daten $[0, t_{max}]$ hinaus wird dadurch vermieden.

Im Kovariablenmodell ist $\tilde{t}_{best}|x$ bei entsprechendem $t_{max}|x$ analog zu erhalten.

Für Verlustfunktionen allgemeinerer Form muß der beste Prädiktor numerisch bestimmt werden. Sind Zensierungen vorhanden, kann $E(L(T, \tilde{t}), \bar{S}(\cdot))$ nicht so weit wie im Fall ohne Zensierungen vereinfacht werden. Der erwartete Fehler mit der Kaplan-Meier(=KM)- Überlebensfunktion $\bar{S}(t)$ würde lauten

$$E(L(T, \tilde{t}), \bar{S}(\cdot)) = \sum_{t_i} L(t_i, \tilde{t}) [\bar{S}(t_i^-) - \bar{S}(t_i)]$$

Hierbei ist für ein zensiertes Element $\bar{S}(t_i^-) - \bar{S}(t_i)$ Null. Das heißt, zensierte Beobachtungen werden bei der Vorhersage nicht berücksichtigt. Eine geeignetere Formulierung verwendet deshalb nur die nichtzensierten Elemente j und lautet mit $t_1, \dots, t_j, \dots, t_m$ den nichtzensierten Beobachtungszeiten (Korn&Simon91)

$$E(L(T, \tilde{t}), \bar{S}(\cdot)) = \sum_{t_j} L(t_j, \tilde{t}) [\bar{S}(t_j^-) - \bar{S}(t_j)]$$

Das Problem ergibt sich im kategorialen Kovariablenmodell analog. Sind Zensierungen vorhanden, ist also im Falle allgemeiner Verlustfunktionen eine numerische Lösung nicht unmittelbar möglich. Eine parametrische Lösung oder eine andere Alternative ist angezeigt.

Binäre Verlustfunktionen

Bei Verwendung der quadratischen binären Verlustfunktion tritt das gleiche Problem bei allgemeinen Verlustfunktionen auf. Ist bei der Berechnung von $\frac{1}{n} \sum_{t_i} I_{\{t_i > t_b\}}$ die Beobachtungszeit eines zensierten Elementes kleiner als der interessierende Zeitpunkt, $t_i < t_b$, kann der beobachtete Status nicht angegeben werden. Eine Schätzung der Überlebenswahrscheinlichkeit beziehungsweise des Status zum Zeitpunkt t_b ist also wie oben nicht direkt als numerische Lösung, sondern nur aufgrund der empirischen Verteilungsfunktion möglich. Für das Kovariablenmodell etwa gilt $\tilde{S}(t_b|x)_{best} = \tilde{S}_{best}(t_b|x) = \bar{S}(t_b|x)$.

Parametrische und semiparametrische Verlustfunktionen

Die Verwendung von $\frac{1}{n}$ als empirische Verteilung von T ist bei parametrischen und semiparametrischen Verlustfunktionen auch im Falle auftretender Zensierungen korrekt, da der Fehler dieser Elemente in der Likelihood geeignet berücksichtigt wird.

Für $-\frac{1}{n}pl(\tilde{\beta})$ bleiben wegen δ_i im Exponenten der einzelnen Komponenten der PL im Falle von Zensierungen nur die Komponenten der nichtzensierten Elemente j .

3.3 Parametrische Schätzung der Vorhersage

Ein Schätzer des erwarteten Fehlers bei nicht vollständig bekannter Verteilung lautetet (Korn&Simon90)

$$E(L(T, \tilde{t}), S(\cdot|\hat{\theta})) := E(L(T, \tilde{t}), \hat{S}(\cdot)) = \int_0^{\infty} L(t, \tilde{t}) d(1 - \hat{S}(t))$$

Gleiches gilt für das Kovariablenmodell mit $S(\cdot|\hat{\theta}, x) =: \hat{S}(\cdot|x)$.

Es ist zunächst $\hat{\theta}$ zum Beispiel als ML -Schätzer zu bestimmen. Dann kann der beste Prädiktor laut Verlustfunktion aus der Verteilung oder für allgemeine Verlustfunktionen numerisch aus dem erwarteten Verlust bestimmt werden.

Sind bei Verwendung des Cox-Modells aufgrund der Verlustfunktion Größen als Vorhersage gesucht, die sich aus der Verteilung ableiten lassen, wie etwa Erwartungswert oder Median, kann nach der Schätzung der Effekte mit Hilfe der PL der Bresow-Schätzer $\hat{S}_0(t|\hat{\beta})$ der Baseline-Überlebensfunktion bestimmt werden um eine vollständige Verteilung zu erhalten.

Es wird hier bemerkt, daß bei der parametrischen Schätzung der Vorhersage davon ausgegangen wird, daß in der Grundgesamtheit keine Zensierungen auftreten, und daß somit bei der Vorhersage entsprechend geglättet und immer extrapoliert wird.

Bei quadratischer Verlustfunktion und geschätzter Normalverteilung folgt der Erwartungswert der geschätzten Verteilung als bester Prädiktor. Dieser ist bei der Normalverteilung identisch mit dem Parameter μ , also $\tilde{t}_{best} = E(T, F(\cdot|\hat{\mu}, \sigma^2)) = \hat{\mu}$. Der erwartete quadratische Fehler lautet

$$E((T - \tilde{t})^2, F(\cdot|\hat{\mu}, \sigma^2)) = \int_0^{\infty} (t - \tilde{t})^2 d\hat{F}(t)$$

, woraus ebenfalls folgt $\tilde{t}_{best} = \hat{\mu}$.

Bei der Exponentialverteilung ist der geschätzte Parameter $\hat{\lambda}$. Der Mittelwert der geschätzten Verteilung als bester Prädiktor läßt sich wiederum als Funktion von dem geschätzten Parameter bestimmen, $\tilde{t}_{best} = E(T, F(\cdot|\hat{\lambda})) = \frac{1}{\hat{\lambda}}$.

Binäre Verlustfunktionen

Bei Verwendung der quadratischen binären Verlustfunktion ist analog zu oben

$$\int_0^\infty L(t, \tilde{S}(t_b)) d(1 - \hat{S}(t)) = \int_0^\infty (I_{\{t > t_b\}} - \tilde{S}(t_b))^2 d(1 - \hat{S}(t))$$

zu minimieren, was zu $\tilde{S}_{best}(t_b) = E(I_{\{T > t_b\}}, \hat{S}(\cdot)) = \hat{S}(t_b)$ führt.

Für die integrierte quadratische binäre Verlustfunktion folgt entsprechend $\tilde{S}_{best}(\cdot) = \hat{S}(\cdot)$.

Parametrische und semiparametrische Verlustfunktionen

Lautet die Verlustfunktion $-l(\theta|t)$, ist also die Vorhersage bestimmter Parameter innerhalb einer gewünschten Verteilungsklasse ausreichend, soll diese Verteilungsklasse im Weiteren mit der Verteilungsklasse der geschätzten Verteilung übereinstimmen. Es ergibt sich wegen $\tilde{\theta}_{opt} = \theta$ für $\tilde{\theta}_{best} = \hat{\theta}$. Der erwartete Fehler lautet

$$E(-l(\tilde{\theta}|T), \hat{F}(\cdot)) = \int_0^\infty -\ln f(t|\tilde{\theta}) f(t|\hat{\theta}) dt$$

beziehungsweise im parametrischen Kovariablenmodell

$$E_x E(-l(\tilde{\theta}|T, x), \hat{F}(\cdot|x)) = \frac{1}{n} \sum_{x_i} \int_0^\infty -\ln f(t|\tilde{\theta}, x_i) f(t|\hat{\theta}, x_i) dt$$

Sind nur die Effekte der Kovariablen in einer nur semiparametrischen Verlustfunktion vorherzusagen, ist deren Erwartungswert zu minimieren. Da hierzu jedoch weder die volle allgemeine *PH*-Likelihood, noch die *PL* geeignet sind (siehe "Parametrische Verlustfunktionen"), kann dieser Erwartungswert zur Herleitung von $\tilde{\theta}_{best} = \hat{\theta}$ beziehungsweise hier nur $\tilde{\beta}_{best} = \hat{\beta}$, nicht genutzt werden.

4 Erklärte Variation

Der Vorhersagewert eines Modells hängt von verschiedenen Faktoren ab. Zunächst von der Modellwahl und – wenn das beste Modell gewählt wurde – inwieweit dessen Modellannahmen erfüllt sind. Weiterhin beeinflusst die Stichprobenvarianz die Varianz der geschätzten Parameter und somit der Vorhersage.

Ein weiterer Faktor ist der Erklärungsgrad der Kovariablen. Er spiegelt sich in der Größe und der Varianz der geschätzten Effekte der Kovariablen wider. Die Signifikanz eines Effekts allein kann auf einem sehr großen Stichprobenumfang beruhen.

Außerdem spielt die Verteilung der Kovariablen für den Erklärungsgrad der Kovariablen eine Rolle; trotz korrektem Modell, niedriger Stichprobenvarianz bei mäßigem Stichprobenumfang und signifikanter Effekte, kann sich eine niedrige Präzision der Vorhersagen ergeben, zum Beispiel bei Verwendung binärer Kovariablen zur Vorhersage einer quantitativen Zielgröße. Ebenfalls von Bedeutung ist die Skalierung der Zielgröße. So kann eine binäre Zielgröße auch bei perfekt trennendem Modell selten mit vollständiger Präzision vorhergesagt werden.

Ein verteilungsfreier Schätzer für den Anteil der durch ein Kovariablenmodell erklärten Varianz an der Gesamtvarianz im Falle quadratischer Verlustfunktion, R^2 , ist der multiple Korrelationskoeffizient r^2 . Allgemein soll als Maß für den Vorhersagewert eines Kovariablenmodells der Anteil des durch das Kovariablenmodell erklärten Fehlers am gesamten Fehler geschätzt werden. Der "gesamte Fehler" ist bei der Schätzung genauer zu spezifizieren.

Bei gleichzeitiger Betrachtung von T und X sind die bedingten Verteilungen $F(t|x)$ als wahre Verteilung anzusehen, aus der sich die Randverteilung für das Modell ohne Kovariablen, $F_0(t)$, ergibt. Der erwartete Fehler für das Kovariablenmodell ist dann der Erwartungswert des erwarteten Fehlers, gegeben x , gemittelt über die Verteilung von X , $G(x)$

$$\int_{-\infty}^{\infty} \int_0^{\infty} L(t, \tilde{t}_{opt}|x) dF(t|x) dG(x)$$

Sollen die Kovariablen bei der Vorhersage nicht verwendet werden, ergibt sich der Prädiktor theoretisch aus der Minimierung des folgenden erwarteten Fehlers

$$\min_{\tilde{t}} \int_{-\infty}^{\infty} \int_0^{\infty} L(t, \tilde{t}) dF(t|x) dG(x) \quad \text{unter der Bedingung} \quad \tilde{t}_{opt}|x = \tilde{t}_{opt} \Rightarrow \tilde{t}_{0_{opt}}$$

Alternativ kann die Vorhersage ohne Verwendung der Kovariablen gefolgert werden aus

$$\min_{\tilde{t}} \int_0^{\infty} L(t, \tilde{t}) d \left(\int_{-\infty}^{\infty} F(t|x) dG(x) \right) \Rightarrow \tilde{t}_{0_{opt}}$$

Die verwendete gewichtete Mischung der bedingten Verteilungen des Kovariablenmodells wird auch Verteilung des Nullmodells genannt (Korn&Simon90)

$$F_0(t) := E_x[F(t|x)] := \int_{-\infty}^{\infty} F(t|x) dG(x)$$

Im Allgemeinen entspricht deren Verteilungsklasse nicht der Verteilungsklasse des Kovariablenmodells; eine Normalverteilungsmodell für das Kovariablenmodell zieht keine Normalverteilung für das Nullmodell nach sich, sondern eine Mischung von Normalverteilungen mit variierenden Mittelwerten.

Der Anteil des durch das Kovariablenmodell erklärten Fehlers an dem Fehler bei Verwendung des gleichen Modells, wenn die Kovariablen nicht verwendet werden, heißt erklärte Variation

$$EV := 1 - \frac{E_x[\int_0^{\infty} L(t, \tilde{t}_{opt}|x) dF(t|x)]}{E_x[\int_0^{\infty} L(t, \tilde{t}_{0opt}) dF(t|x)]}$$

beziehungsweise alternativ

$$EV_{F_0} := 1 - \frac{E_x[\int_0^{\infty} L(t, \tilde{t}_{opt}|x) dF(t|x)]}{\int_0^{\infty} L(t, \tilde{t}_{0opt}) dF_0(t)}$$

Diese gibt den Erklärungsgrad der Kovariablen für die Zielgröße an, der bei optimaler Vorhersage in dem wahren Modell, also maximal, erreicht werden kann.

Für den quadratischen Fehler folgt

$$R_{F_0}^2 := 1 - \frac{E_x[Var(T, F(\cdot|x))]}{Var(T, F_0(\cdot))}$$

Die EV mit Werten zwischen 0 und 1 ist eine Ergänzung bei der Beurteilung eines Vorhersagemodells. So ist nicht nur eine Beurteilung möglich, wie gut die Kovariablen die Ausprägungen der Zielgröße trennen, sondern mit welcher Präzision sie sie quantitativ vorhersagen.

Binäre Verlustfunktionen

Bei Verwendung einer binären Verlustfunktion $L(t, \tilde{S}(t_b))$ mit vorherzusagendem

Status zu einem bestimmten Zeitpunkt t_b bei realisierter Ausfallzeit t gilt für $S(\tilde{t}_b)_{opt}|x = \tilde{S}_{opt}(t_b|x)$ und für $S(\tilde{t}_b)_{0_{opt}} = \tilde{S}_{0_{opt}}(t_b)$. Die EV ergibt sich bei quadratischer Verlustfunktion dann wie oben durch Ersetzung von \tilde{t}_{opt} durch $\tilde{S}_{opt}(t_b) = S(t_b)$ und t durch $I_{\{t>t_b\}}$

$$\begin{aligned} R_{F_0,bin}^2 &:= 1 - \frac{E_x[\int_0^\infty (I_{\{t>t_b\}} - S(t_b|x))^2 dF(t|x)]}{\int_0^\infty (I_{\{t>t_b\}} - S_0(t_b))^2 dF_0(t)} \\ &= 1 - \frac{E_x[S(t_b|x)(1 - S(t_b|x))]}{S_0(t_b)(1 - S_0(t_b))} \end{aligned}$$

Ist die EV eines binären Prädiktors für den gesamten Zeitverlauf von Interesse, ergibt sich durch Integration der Verlustfunktion über t_b und Verwendung der integrierten quadratischen binären Fehler in EV

$$\begin{aligned} R_{F_0,intbin}^2 &:= 1 - \frac{E_x[\int_0^\infty \int_0^\infty (I_{\{t>t_b\}} - S(t_b|x))^2 dt_b dF(t|x)]}{\int_0^\infty \int_0^\infty (I_{\{t>t_b\}} - S_0(t_b))^2 dt_b dF_0(t)} \\ &= 1 - \frac{E_x[\int_0^\infty S(t_b|x)(1 - S(t_b|x)) dt_b]}{\int_0^\infty S_0(t_b)(1 - S_0(t_b)) dt_b} \end{aligned}$$

Parametrische und semiparametrische Verlustfunktionen

Bei Verwendung der negativen log-Likelihood-Funktion als Verlustfunktion, $-l(\tilde{\theta}|t, x)$, in einem parametrischen Kovariablenmodell mit $\tilde{\theta} = (\tilde{\theta} \setminus \tilde{\beta}, \tilde{\beta})$ ist der Parameter im Falle des Modells ohne Verwendung der Kovariablen definiert durch

$$\min_{\tilde{\theta}} -F_c(\tilde{\theta}) = \min_{\tilde{\theta}} \int_{-\infty}^\infty \int_0^\infty -\ln f(t|\tilde{\theta}, x) f(t|x) dt dG(x) \quad \text{bei} \quad \tilde{\beta}_{opt} = 0 \quad \Rightarrow \quad \tilde{\theta}_{opt}$$

Eine Vereinfachung bei der Bestimmung der Vorhersage mit Hilfe der Nullverteilung ist hier nicht möglich, da x in der Verlustfunktion enthalten ist.

Zunächst scheint es naheliegend, die EV als Entropie-basiertes Maß in Form des Verhältnisses der erwarteten negativen log-Likelihood-Funktionen von Kovariablenmodell und Modell ohne Verwendung der Kovariablen zu definieren

$$EV_{Lik} := 1 - \frac{-F_c(\tilde{\theta}_{opt})}{-F_c(\tilde{\theta}_{0_{opt}})}$$

Die Motivation für einen anderen Ansatz ergibt sich wegen

$$\exp[2 - F(\tilde{\mu}_{opt}, \tilde{\sigma}_{opt}^2)] = \exp[E(-2l(\tilde{\mu}_{opt}|T), F(\cdot))] = Var(T, F(\cdot)) * 2\pi * e$$

im Falle der Normalverteilung. Es ist auch im allgemeinen Fall sinnvoll, die exponentierte erwartete negative log-Likelihood-Funktion in der EV zu verwenden (siehe auch "Parametrische Verlustfunktionen"). Dann ergibt sich

$$\begin{aligned} EV_{LR} &:= 1 - \frac{\exp[2 - F_c(\tilde{\theta}_{opt})]}{\exp[2 - F_c(\tilde{\theta}_{0_{opt}})]} \\ &= 1 - \frac{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|\tilde{\theta}_{opt}, x) f(t|x) dt\right]\right)}{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|\tilde{\theta}_{0_{opt}}) f(t|x) dt\right]\right)} \end{aligned}$$

Bei der Bestimmung des Fehlers im Nenner kann nun

$$\exp\left(\int_0^\infty -2 \ln f(t|\tilde{\theta}_{0_{opt}}) f_0(t) dt\right)$$

verwendet werden.

Die ursprüngliche Definition beruht auf einer Transformation des Kullback-Leibler- Informationsgewinns

$$\Gamma(\tilde{\theta}_{opt}, \tilde{\theta}_{0_{opt}}) := 2 \left(F_c(\tilde{\theta}_{opt}) - F_c(\tilde{\theta}_{0_{opt}}) \right)$$

Es ist (Kent83)

$$\begin{aligned} EV_{LR} &= 1 - \exp(-2(F_c(\tilde{\theta}_{opt}) - F_c(\tilde{\theta}_{0_{opt}}))) \\ &= 1 - \exp(-\Gamma(\tilde{\theta}_{opt}, \tilde{\theta}_{0_{opt}})) \end{aligned}$$

mit den Eigenschaften (Kent&O'Quigley88, Nagelkerke91):

- Bei Normalverteilung ergibt sich R^2 .
- Konsistenz bezüglich der ML -Theorie:
 EV_{LR} wird bei $f(t|x) = f(t|\theta, x)$ maximal für θ .

- Normierung für diskret verteilte T :

Für Modelle, deren Likelihood eine Wahrscheinlichkeit darstellt und da diese stets positiv ist, gilt $0 \geq F_c(\tilde{\theta}_{opt}) \geq F_c(\tilde{\theta}_{0_{opt}})$ und $\max(EV_{LR}) = 1 - \exp(2F_c(\tilde{\theta}_{0_{opt}})) < 1$.

Man verwende in diesem Fall

$$EV_{LR,norm} := \frac{EV_{LR}}{\max(EV_{LR})}$$

Semiparametrische Verlustfunktionen

Um eine EV unter Verwendung einer semiparametrischen Verlustfunktion zu definieren, die das Cox-Modell repräsentiert, kann der in "Parametrische Verlustfunktionen" beschriebene Ansatz von Kent&O'Quigley88 verwendet werden. Um eine vollständige Dichte für die Formulierung der erwarteten log-Likelihood in geschlossener Form zu erhalten, wird das Cox-Modell zu einem Weibullverteilungsmodell spezifiziert. Hierbei werden die Effekte β aus dem Cox-Modell und zusätzlich beliebige Werte der fehlenden Parameter des Weibullmodells, α^* und a^* , von extern in eine Weibullverteilung insgesamt als θ^* eingebracht, um eine vollständige wahre Verteilung anzugeben. Betrachtet man die EV als Maßzahl zum Vergleich zweier Hypothesen, so ist diese Ergänzung zu einer fiktiven Verteilung unbedingt als Teil beider Hypothesen zu betrachten, um die Unabhängigkeit der EV von α^* und a^* zu gewährleisten (Kent&O'Quigley88).

$$\begin{aligned} EV_{PLR^*} &:= 1 - \frac{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|\tilde{\theta}_{opt}^*, x) f(t|\theta^*, x) dt\right]\right)}{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|\tilde{\theta}_{0_{opt}}^*) f(t|\theta^*, x) dt\right]\right)} \\ &= 1 - \frac{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|(\beta, \alpha^*, a^*), x) f(t|(\beta, \alpha^*, a^*), x) dt\right]\right)}{\exp\left(E_x\left[\int_0^\infty -2 \ln f(t|(0, \tilde{\alpha}_{0_{opt}}^*, \tilde{a}_{0_{opt}}^*)) f(t|(\beta, \alpha^*, a^*), x) dt\right]\right)} \end{aligned}$$

Die so formulierte EV ist invariant gegenüber linearen Transformationen der Zeit und somit insbesondere gegenüber den zusätzlich für die Spezifikation des Weibullmodells zu wählenden Parametern.

So können die gegenüber monotonen Transformationen der Zeit invarianten Kovariableneffekte beurteilt werden, die in einem Cox-Modell verwendet werden.

5 Schätzung der erklärten Variation

Bei der Schätzung der EV wird wieder wie bei der Schätzung der Vorhersage von festen Ausprägungen der Kovariablen ausgegangen und als deren Verteilung $G(x)$ die empirische Verteilung von X verwendet.

5.1 Schätzung der erklärten Variation unter Modellannahme

Wurden die Vorhersagen parametrisch geschätzt, das heißt ist eine parametrisch geschätzte Verteilung für die zukünftigen Beobachtungen vorhanden, $S(t|\hat{\theta}, x) = \hat{S}(t|x)$, kann diese verwendet werden, um unter der Annahme, daß dieses Modell das korrekte Modell darstellt, den erwarteten Fehler zu berechnen, sowohl für das Kovariablenmodell

$$\frac{1}{n} \sum_{x_i} E(L(T, \tilde{t}_{best}|x), \hat{S}(\cdot|x)) = \frac{1}{n} \sum_{x_i} \int_0^{\infty} L(t, \tilde{t}_{best}|x_i) d(1 - \hat{S}(t|x_i))$$

als auch für das Nullmodell

$$E(L(T, \tilde{t}_{0_{best}}|x), \hat{S}_0(\cdot)) = \int_0^{\infty} L(t, \tilde{t}_{0_{best}}) d(1 - \hat{S}_0(t))$$

mit

$$\hat{S}_0(t) = \frac{1}{n} \sum_{x_i} \hat{S}(t|x_i)$$

Anschließend kann

$$\hat{EV}_{F_0} := 1 - \frac{\frac{1}{n} \sum_{x_i} \int_0^{\infty} L(t, \tilde{t}_{best}|x_i) d(1 - \hat{S}(t|x_i))}{\int_0^{\infty} L(t, \tilde{t}_{0_{best}}) d(1 - \hat{S}_0(t))}$$

angegeben werden (Korn&Simon90). Zum Beispiel bei quadratischer Verlustfunktion ist

$$\hat{R}_{F_0}^2 := 1 - \frac{\frac{1}{n} \sum_{x_i} Var(T, \hat{F}(\cdot, x_i))}{Var(T, \hat{F}_0(\cdot))}$$

Bei dieser Methode werden die Daten, die zur Modellbildung verwendet wurden, nicht auch zur Modellbeurteilung verwendet. Dadurch werden die tatsächlich aufgetretenen Verluste nach dem gebildeten Modell geglättet und extrapoliert. In der

Anwendung bedeutet dies, daß das Problem der Beurteilung der Verluste der Vorhersage für zensierte Daten fort fällt.

Allerdings hängt die Aussagekraft von $\hat{E}V_{F_0}$ von dem Grad der Validität des Modells ab. Hypothetisch können außerdem Modelle, die aufgrund unterschiedlich beschaffener Datensätze gewonnen wurden (zum Beispiel tatsächlich größere Varianz, jedoch ausgleichend größerer Stichprobenumfang) zu identischen Ergebnissen bezüglich des Erklärungsgrades der Kovariablen führen.

Binäre Verlustfunktionen

Bei Verwendung binärer Verlustfunktionen ist ebenso vorzugehen, wodurch sich in diesem Falle bei quadratischer Verlustfunktion ergibt

$$\begin{aligned}\hat{R}_{F_0,bin}^2 &= 1 - \frac{\frac{1}{n} \sum_{x_i} \int_0^\infty (I_{\{t>t_b\}} - \hat{S}(t_b|x_i))^2 d\hat{S}(t|x_i)}{\int_0^\infty (I_{\{t>t_b\}} - \hat{S}_0(t_b))^2 d\hat{S}_0(t)} \\ &= 1 - \frac{\frac{1}{n} \sum_{x_i} \hat{S}(t_b|x_i)(1 - \hat{S}(t_b|x_i))}{\hat{S}_0(t_b)(1 - \hat{S}_0(t_b))}\end{aligned}$$

und für die integrierte quadratische binäre Verlustfunktion

$$\hat{R}_{F_0,intbin}^2 = 1 - \frac{\frac{1}{n} \sum_{x_i} \int_0^\infty \hat{S}(t_b|x_i)(1 - \hat{S}(t_b|x_i)) dt_b}{\int_0^\infty \hat{S}_0(t_b)(1 - \hat{S}_0(t_b)) dt_b}$$

Parametrische und semiparametrische Verlustfunktionen

Wie unter "Parametrische Schätzung der Vorhersage" wird auch hier davon ausgegangen, daß die Verteilungsklasse der log-Likelihood und der geschätzten Verteilung übereinstimmen. Es ergibt sich als Prädiktor der Schätzer, $\tilde{\theta}_{best} = \hat{\theta}$, und für die erklärte Variation (Kent&O'Quigley88)

$$\begin{aligned}\hat{E}V_{LR} &:= 1 - \frac{\exp\left(\frac{1}{n} \sum_{x_i} \int_0^\infty -2 \ln f(t|\hat{\theta}, x_i) f(t|\hat{\theta}, x_i) dt\right)}{\exp\left(\frac{1}{n} \sum_{x_i} \int_0^\infty -2 \ln f(t|\hat{\theta}_0) f(t|\hat{\theta}, x_i) dt\right)} \\ &= 1 - \frac{\exp[-2F_c(\hat{\theta})]}{\exp[-2F_c(\hat{\theta}_0)]} \\ &= 1 - \exp(-\Gamma(\hat{\theta}, \hat{\theta}_0))\end{aligned}$$

Unter der Annahme eines Cox-Modells wird wie bei der Definition der EV_{PLR^*} auch bei deren Schätzung \hat{EV}_{PLR^*} eine Spezifizierung des Cox-Modells auf ein Weibullmodell verwendet (siehe "Parametrische Verlustfunktionen"), um die entsprechenden geschätzten erwarteten log-Likelihood-Funktionen zu formulieren. Hierbei sind die mit Hilfe der PL geschätzten Effekte $\hat{\beta}$ und beliebig zu wählende Parameter des Weibullmodells, etwa $\hat{\alpha}^* = 0$ und $\hat{a}^* = 1$, insgesamt $\hat{\theta}^*$, für die Schätzung des erwarteten Fehlers des Kovariablenmodells zu verwenden. Für das Modell ohne Verwendung der Kovariablen ergeben sich aus der geschätzten erwarteten log-Likelihood unter der Bedingung $\hat{\beta}_0 = 0$ die Parameter $\hat{\alpha}_0^*$ und \hat{a}_0^* , insgesamt $\hat{\theta}_0^*$. Der Wert von \hat{EV}_{PLR^*} ist wieder unabhängig von der Wahl der zusätzlich zu wählenden Parameter des Weibullmodells.

$$\hat{EV}_{PLR^*} := 1 - \frac{\exp\left(\frac{1}{n} \sum_{x_i} \int_0^\infty -2 \ln f(t|(\hat{\beta}, 0, 1), x) f(t|(\hat{\beta}, 0, 1), x)\right)}{\exp\left(\frac{1}{n} \sum_{x_i} \int_0^\infty -2 \ln f(t|(0, \hat{\alpha}_0^*, \hat{a}_0^*)) f(t|(\hat{\beta}, 0, 1), x)\right)}$$

5.2 Schätzung der erklärten Variation ohne Modellannahme

Die EV kann auch ohne die Annahme, daß das geschätzte Modell das wahre Modell darstellt, geschätzt werden. Es werden die Fehler der Vorhersage nicht unter Verwendung der geschätzten Verteilung, sondern unter Verwendung der empirischen Verteilung von T berechnet, also ohne Glättung und Extrapolation nach dem angenommenen Modell. Der empirische Schätzer der erklärten Variation lautet wegen $\bar{F}_0 = \bar{F}$

$$ev := 1 - \frac{\sum_x E(L(T, \tilde{t}_{best}|x), \bar{S}(\cdot|x)) \frac{n_x}{n}}{E(L(T, \tilde{t}_{0_{best}}), \bar{S}(\cdot))}$$

Bei einem anderen Zugang bezieht sich die Beurteilung auf die Daten. Die vorhergesagten Werte werden direkt den beobachteten Werten gegenübergestellt, es werden die Residuen betrachtet und entsprechend der empirischen Verteilung

gemittelt

$$rv := 1 - \frac{\frac{1}{n} \sum_{t_i | x_i=x} L(t_i, \tilde{t}_{best} | x_i)}{\frac{1}{n} \sum_{t_i} L(t_i, \tilde{t}_{0_{best}})}$$

\tilde{t}_{best} bezeichnet den parametrisch oder verteilungsfrei geschätzten Prädiktor. Es gilt ohne Zensierungen

$$ev = rv = 1 - \frac{\sum_{t_i, x_i} L(t_i, \tilde{t}_{best} | x_i) \frac{1}{n}}{\sum_{t_i} L(t_i, \tilde{t}_{0_{best}}) \frac{1}{n}}$$

Unter Verwendung der quadratischen Verlustfunktion ergibt sich in beiden Fällen für den Nenner die empirische Varianz und für den Zähler die Residuenvarianz. Bei verteilungsfreier Schätzung der Vorhersage mit $\tilde{t}_{best} = \bar{t}$ folgt r^2 .

Binäre Verlustfunktionen

Wird die quadratische binäre Verlustfunktion $L(t, S(\tilde{t}_b)) = (I_{\{t > t_b\}} - S(\tilde{t}_b))^2$ zur Schätzung des Vorhersagewertes einer Überlebenswahrscheinlichkeit $S(\tilde{t}_b)_{best}$ verwendet, ergibt sich ev_{bin} analog wie ev .

Bei Verwendung der integrierten quadratischen binären Verlustfunktion zur Schätzung des Vorhersagewertes der Überlebensfunktion kann statt der Lebesgue-Integration dt_b die empirische Verteilung von T zu einer gewichteten Integration verwendet werden, $d\bar{S}(t_b)$ (Graf&Schumacher95). Dies scheint bei verteilungsfreier Schätzung der EV zur Vermeidung einer Inter- beziehungsweise Extrapolation sinnvoll. Es ergibt sich für den Verlust des einzelnen Elements eine gewichtete Summe nur über tatsächlich beobachtete Zeitpunkte der binären Fehler, also insgesamt statt

$$\frac{1}{n} \sum_{t_i} \left(\int_0^\infty (I_{\{t_i > t_b\}} - \tilde{S}_{best}(t_b))^2 dt_b \right)$$

jetzt

$$\frac{1}{n} \sum_{t_i} \left(\sum_{t_b} (I_{\{t_i > t_b\}} - \tilde{S}_{best}(t_b))^2 \frac{1}{n} \right)$$

Die Summe ergibt als maximalen Wert n . Die Multiplikation mit $\frac{1}{n}$ kann ebenso als Normierung des Fehlers auf ein Maximum von 1 für jedes Element i betrachtet

werden.

Für ein parametrisch geschätztes Kovariablenmodell folgt entsprechend statt

$$\frac{1}{n} \sum_{t_i, x_i} \left(\int_0^\infty (I_{\{t_i > t_b\}} - \hat{S}(t_b|x_i))^2 dt_b \right)$$

jetzt

$$\frac{1}{n} \sum_{t_i, x_i} \left(\sum_{t_b} (I_{\{t_i > t_b\}} - \hat{S}(t_b|x_i))^2 \frac{1}{n} \right)$$

Es wird wieder zur Integration $\bar{S}(t_b)$, nicht $\bar{S}(t_b|x)$, verwendet.

Bei empirisch geschätztem Kovariablenmodell ist es sinnvoll, auch innerhalb der jeweiligen Kategorie x nicht zu inter- oder extrapolieren. Deshalb wird eine Modifikation der Lösung für parametrisch geschätztes Kovariablenmodell vorgeschlagen. Es sollen pro Kategorie x nur die Zeitpunkte $t_i|x_i = x$ für die Verluste verwendet werden, da im nonparametrischen Kovariablenmodell nur auf diesen die Vorhersage beruht. Es wird eine gleiche Wahrscheinlichkeitsmasse für Elemente i aus verschiedenen Kategorien gewährleistet, wenn statt $d\bar{S}(t_b)$ jetzt $d\bar{S}(t_b|x)$ für die Integration verwendet wird, also statt

$$\frac{1}{n} \sum_x \sum_{t_i|x_i=x} \left(\int_0^\infty (I_{\{t_i > t_b\}} - \bar{S}(t_b|x))^2 dt_b \right) = \frac{1}{n} \sum_x \sum_{t_i|x_i=x} \left(\int_0^{t_{max}|x} (I_{\{t_i > t_b\}} - \bar{S}(t_b|x))^2 dt_b \right)$$

jetzt

$$\frac{1}{n} \sum_x \sum_{t_i|x_i=x} \left(\sum_{t_b|x_b=x} (I_{\{t_i > t_b\}} - \bar{S}(t_b|x))^2 \frac{1}{n_x} \right)$$

Als ev_{intbin} folgt generell die Differenz zwischen 1 und dem Verhältnis des Fehlers von Kovariablenmodell und Modell ohne Kovariablen.

Parametrische und semiparametrische Verlustfunktionen

Wird bei der $\hat{E}V_{LR}$ die Modellannahme nicht zur Integration verwendet, ergibt sich die naheliegende Verallgemeinerung des multiplen Korrelationskoeffizienten r^2 für Likelihood-basierte Modelle (Kent83, Maddala83, Cox&Snell89, Nagelkerke91)

$$ev_{LR} := 1 - \frac{\exp[2 \sum_{t_i, x_i} -\ln f(t_i|\hat{\theta}, x_i) \frac{1}{n}]}{\exp[2 \sum_{t_i} -\ln f(t_i|\hat{\theta}_0) \frac{1}{n}]}$$

$$\begin{aligned}
&= 1 - \frac{\exp[2\frac{1}{n} - l(\hat{\theta})]}{\exp[2\frac{1}{n} - l(\hat{\theta}_0)]} \\
&= 1 - \exp(-\hat{\Gamma}(\hat{\theta}, \hat{\theta}_0)) \\
&= 1 - \exp(-\frac{1}{n}LR)
\end{aligned}$$

mit LR der Likelihood-Ratio-Statistik.

Ist in der \hat{EV}_{LR} $\hat{\theta}$ ML -Schätzer und ist f Exponentialfamilie, so sind \hat{EV}_{LR} und ev_{LR} identisch (Kent86).

Es gilt (ähnlich wie unter "Erklärte Variation"):

- Konsistenz bezüglich der ML -Methode; der ML -Schätzer maximiert ev_{LR} .
- Asymptotische Unabhängigkeit von der Stichprobengröße.
- Das Maß ist für stetige Modelle normiert und kann für diskrete Modelle mit $\max(ev_{LR}) = 1 - \exp(2\frac{1}{n}l(\hat{\theta}_0)) < 1$ einfach normiert werden

$$ev_{LR,norm} := \frac{ev_{LR}}{\max(ev_{LR})}$$

- Für Normalverteilung ergibt sich das klassische r^2 der linearen Regression.

Ein weiterer Vorschlag schätzt den Vorhersagewert wie $ev_{LR,norm}$ nach einem Normierungsprinzip, jedoch direkt für die LR . Der maximal erreichbare Wert der LR ergibt sich bei Verwendung des saturierten Modells, $l(sat) := l(\hat{\theta}_1, \dots, \hat{\theta}_n)$. Dieser lautet $\max(LR) = 2(l(sat) - l(\hat{\theta}_0))$. Es ergibt sich dann als alternativer Schätzwert (Harrell97)

$$LR_{norm} := \frac{LR}{\max(LR)} = \frac{l(\hat{\theta}) - l(\hat{\theta}_0)}{l(sat) - l(\hat{\theta}_0)}$$

Im Falle der logistischen Regression gilt $l(sat) = 0$, wodurch sich in diesem Fall wiederum ev_{Lik} (erstmal Thiel70) ergibt

$$ev_{Lik} := 1 - \frac{-l(\hat{\theta})}{-l(\hat{\theta}_0)} = \frac{l(\hat{\theta}) - l(\hat{\theta}_0)}{-l(\hat{\theta}_0)}$$

Deren Zugang basierte jedoch statt auf der Differenz der log-Likelihood-Funktionen wie in LR_{norm} auf dem Verhältnis der log-Likelihood-Funktionen. Im Falle der

logistischen Regression liefert $ev_{Lik} = LR_{norm}$ in der Anwendung plausible Ergebnisse (siehe etwa Mittelböck&Schemper96).

Im Gegensatz zu LR_{norm} lautet die Devianz, die ebenfalls die Likelihood des saturierten Modells verwendet (McCullagh&Nelder83)

$$Dev_{Lik} := 2(l(sat) - l(\hat{\theta})) = \max(LR) - LR$$

Semiparametrische Verlustfunktionen

\hat{EV}_{PLR^*} ist ohne Verwendung der Modellannahme zur Integration ist nicht anzugeben. Die Vervollständigung der Weibullverteilung durch die beliebig zu wählenden Parameter $\hat{\alpha}^*$ und \hat{a}^* ist fester Bestandteil der zu vergleichenden Hypothesen, kann dann jedoch nicht in der angenommenen Verteilung enthalten sein.

Eine Alternative bildet (Schemper92)

$$ev_{PLR} := 1 - \frac{\exp[2\frac{1}{n} - pl(\hat{\beta})]}{\exp[2\frac{1}{n} - pl(0)]},$$

wobei jedoch unklar ist, welche Größe durch ev_{PLR} gemessen wird, da die PL weder als Verlustfunktion, noch deren Erwartungswert angegeben werden kann (siehe auch "Parametrische Verlustfunktionen" und "Parametrische Schätzung der Vorhersage").

Wie für parametrische Verlustfunktionen kann auch hier analog zu LR_{norm} ein Schätzwert PLR_{norm} angegeben werden, welcher in diesem Fall wegen $pl(sat) = 0$ identisch ist mit ev_{PL} (Harrell86). Dieser Schätzer liefert jedoch um auf Bruchteile reduzierte Werte der anderen Schätzer (siehe etwa Schemper90).

Auch die oben angegebene Devianz wird für die PL als Dev_{PL} formuliert (Therneau&al90).

5.3 Schätzung der erklärten Variation bei Zensierungen

Die EV kann für zensierte Daten verteilungsfrei geschätzt werden, indem wie bei der verteilungsfreien Schätzung die empirischen Verteilungen $\bar{S}(t|x)$ und $\bar{S}(t)$ zur Schätzung der Fehler, – etwa bei quadratischer Verlustfunktion der Varianz der KM -Überlebensfunktion – verwendet werden. Die ev ist ein empirisch geschätzter

Populationsparameter und ändert sich bei variierendem Anteil an Zensierungen nicht. Jedoch sollte davon ausgegangen werden, daß ein höherer Zensierungsanteil bei der Schätzung die Präzision der Vorhersage erniedrigt.

Deshalb ist der Fehler mit Hilfe der rv zu bestimmen wie im Falle allgemeiner Verlustfunktionen. Bei Verwendung der empirischen Verteilung zur Gewichtung bei der Mittelung werden dabei jedoch nur die nicht zensierten Elemente berücksichtigt. Verschiedene Alternativen beziehen die Residuen von zensierten Elementen unter Verwendung bestimmter Annahmen über eine Extrapolation mit ein.

Gewichtung der Kategorien

Für kategoriale X kann bei noninformativem Zensierungsmechanismus bei gleichem Risikoprofil im Mittel der gleiche Verlust für die zensierten Elemente angenommen werden, wie für die nichtzensierten. Es folgt in Anlehnung an *ev* eine Gewichtung der mittleren Fehler der nichtzensierten Elemente in den Kategorien nach dem gesamten Stichprobenumfang der Kategorie, so daß Zensierungen mittelbar eingehen (Korn&Simon91)

$$rv_{kat} := 1 - \frac{\sum_x \left(\sum_{t_j | x_j=x} L(t_j, \tilde{t}_{best}|x) [\bar{S}(t_j^-|x) - \bar{S}(t_j|x)] \right) \frac{n_x}{n}}{\sum_{t_j} L(t_j, \tilde{t}_{best}) [\bar{S}_0(t_j^-) - \bar{S}_0(t_j)]}$$

Es ist \bar{S}_0 statt \bar{S} zu verwenden, da nicht mehr von unvollständigen Beobachtungen bei der Gewichtung der Verluste ausgegangen werden soll. Für stetige Kovariablen ergibt sich für das nicht zensierte Element j mit der individuell beobachteten Überlebensfunktion

$$\bar{S}(t|x_j) = \begin{cases} 1 & : t < t_j \\ 0 & : \text{sonst} \end{cases}$$

und m der Anzahl der nicht zensierten Elemente jetzt für den geschätzten Fehler des Kovariablenmodells

$$\begin{aligned} & \left(\sum_{(t_j, x_j)} L(t_j, \tilde{t}_{best}|x_j) [\bar{S}(t_j^-|x_j) - \bar{S}(t_j|x_j)] \right) \frac{1}{m} \\ &= \left(\sum_{(t_j, x_j)} L(t_j, \tilde{t}_{best}|x_j) [1] \right) \frac{1}{m} \end{aligned}$$

Dieser stützt sich ausschließlich auf die nicht zensierten Elemente.

Konservative Verlustfunktion

Eine andere Möglichkeit, die EV zu schätzen, ohne die Residuen der zensierten Elemente zu vernachlässigen und diese auch bei stetigen X zu berücksichtigen, besteht in der Definition einer konservativen Verlustfunktion (Henderson95)

$$L_{kons}((t_i, \delta_i), \tilde{t}) := \begin{cases} L(t_i, \tilde{t}) & : (\delta_i = 1) \text{ oder } (t_i > \tilde{t}) \\ 0 & : \text{sonst} \end{cases}$$

Eine vorhergesagte Ausfallzeit beliebig größer als die beobachtete Zensierungszeit wird als korrekt vorhergesagt angesehen und eine Ausfallzeit kleiner als die beobachtete Zensierungszeit erhält als Fehler nur die Differenz bis zur Zensierungszeit. Wird der fehlende Verlust für die zensierten Elemente durch den jeweils minimal anzunehmenden Verlust ersetzt, kann die Verwendung der KM -Überlebensfunktion als empirischer Verteilungsfunktion zugunsten der üblichen empirischen Verteilung aufgegeben werden. Der geschätzte Fehler des Kovariablenmodells, der geschätzte Fehler des Nullmodells und schließlich rv_{kons} kann nun formuliert werden.

Multiple Ergänzung

Weiterhin können statt der Verluste für die zensierten Elemente die Beobachtungszeiten der zensierten Elemente ergänzt werden. Dies kann für rangskalierte Zeit zufällig durch multiple Imputation der beobachteten Zensierungszeiten zu Ausfallzeiten geschehen (Schemper96).

Binäre Verlustfunktionen

Wird die quadratische binäre Verlustfunktion $L(t, S(\tilde{t}_b)) = (I_{\{t > t_b\}} - S(\tilde{t}_b))^2$ zur Schätzung des Vorhersagewertes einer Überlebenswahrscheinlichkeit $S(\tilde{t}_b)_{0_{best}}$ bei zensierten Überlebenszeiten verwendet, ergibt sich das gleiche Problem wie bei der verteilungsfreien Schätzung der Überlebenswahrscheinlichkeit ohne Schätzung der gesamten Überlebensfunktion. Ist die Beobachtungszeit eines zensierten Elementes kleiner als der interessierende Zeitpunkt, $t_i < t_b$, kann der beobachtete Status für die Berechnung des quadratischen binären Fehlers nicht angegeben werden.

Sind die Überlebenswahrscheinlichkeiten für den gesamten Zeitverlauf von Interesse, so kann der integrierte quadratische binäre Fehler für ein nichtzensiertes

Element wie unter "Schätzung der erklärten Variation ohne Modellannahme" geschätzt werden. Wird hierbei bei der Integration die empirische Verteilung verwendet, so werden nur tatsächlich beobachtete Ausfallzeiten t_j statt t_b verwendet und die einzelnen binären Verluste im Zeitverlauf laut der empirischen Verteilung von T gewichtet.

Bei parametrischem Kovariablenmodell folgt

$$\frac{1}{n} \sum_{t_i, x_i} \left(\sum_{t_j} (I_{\{t_i > t_j\}} - \hat{S}(t_j | x_i))^2 [\bar{S}(t_j^-) - \bar{S}(t_j)] \right)$$

Wurde Element i zensiert, so kann für $t_j > t_i$ keine Aussage für den beobachteten Status $I_{\{t_i > t_j\}}$ gemacht werden. Deshalb ist für zensierte Elemente zu verwenden

$$\frac{1}{n} \sum_{t_i, x_i} \left(\sum_{(t_j \leq t_i)} (I_{\{t_i > t_j\}} - \hat{S}(t_j | x_i))^2 [\bar{S}(t_j^-) - \bar{S}(t_j)] \right)$$

Es geht dann für ein zensiertes Element nur der bis zur Zensierung meßbare Fehler mit ein, der gewichtet wird nach der Dauer unter Beobachtung und immer geringer ist als der Verlust eines nichtzensierten Elements, selbst wenn dieses kürzer unter Beobachtung stand.

Bei verteilungsfreier Schätzung der Vorhersage sollten analog wie im Falle ohne Zensierungen für das Kovariablenmodell nur die tatsächlich in der Kategorie beobachteten Ausfallzeitpunkte und die empirischen Verteilungen $\bar{S}(\cdot | x)$ für die Mittelung für ein Element i aus der Kategorie x verwendet werden. Ein Verhindern der Extrapolation des binären Fehlers über die erhobenen Daten hinaus ist bei vorliegenden Zensierungen außerdem deshalb notwendig, da im Falle von Zensierungen am Ende der Beobachtungsdauer in einer Kategorie x für ein nichtzensiertes Element aus dieser Kategorie für $t > t_{max} | x$ der beobachtete Status zwar angegeben werden kann, die vorhergesagte Überlebenswahrscheinlichkeit aber nicht definiert ist.

Ein anderer Vorschlag läßt sich bei Vorliegen von Zensierungen nicht auf die Idee der Integration mit Hilfe der empirischen Verteilung zurückführen, sondern auf die Idee einer Normierung des Verlustes pro Element auf ein Maximum von 1 wie unter "Schätzung der erklärten Variation ohne Modellannahme" beschrieben, so daß anders als beim letzten Vorschlag zensierte Elemente das gleiche Gewicht

bei der Fehlerberechnung erhalten, wie nichtzensierte. Hierzu wird der über alle aufgetretenen Ausfallzeiten t_j aufsummierte Verlust jeweils durch die Anzahl der Summanden geteilt. Für zensierte Elemente werden wie oben nur die Ausfallzeitpunkte $t_j < t_i$ verwendet. Sei die Anzahl der verwendeten Ausfallzeitpunkte für ein Element i

$$m_i := \begin{cases} m & : \delta_i = 1 \\ \#\{j|t_j \leq t_i\} & : \delta_i = 0 \end{cases}$$

Dann geht bei parametrischem Kovariablenmodell der Verlust zensierter Elemente mit dem gleichen Gewicht in den integrierten quadratischen binären Fehler ein wie der Verlust eines nichtzensierten Elements (Schemper90)

$$\frac{1}{n} \sum_{t_i, x_i} \left(\sum_{j=1}^{m_i} (I_{\{t_i > t_j\}} - \tilde{S}(t_j|x_i))^2 \frac{1}{m_i} \right)$$

Es folgt

$$rv_{Sch} := 1 - \frac{\frac{1}{n} \sum_{t_i, x_i} \left(\sum_{j=1}^{m_i} (I_{\{t_i > t_j\}} - \tilde{S}(t_j|x_i))^2 \frac{1}{m_i} \right)}{\frac{1}{n} \sum_{t_i} \left(\sum_{j=1}^m (I_{\{t_i > t_j\}} - \tilde{S}(t_j))^2 \frac{1}{m} \right)}$$

6 Erklärte Variation mit Martingal-Residuen

Martingal für Überlebenszeiten

Eine weitere Möglichkeit zur Schätzung der EV ohne Modellannahme bei Zensierungen besteht in der Formulierung von Residuen mit Hilfe von individuellen Zählprozessen (Fleming&Harrington91).

Sei für ein Element i

$$\delta_i(t) := I_{\{(t_i \leq t) \wedge (\delta_i = 1)\}}$$

der Ausfallprozeß. Da die Zustände "Ausfall" oder "Zensierung" absorbierend sind, wird außerdem der Risikoprozeß

$$r_i(t) := I_{\{t_i \geq t\}}$$

definiert. Es gilt $r(t) := \sum r_i(t) = |R(t)|$.

Dann kann der Intensitätsprozeß in Abhängigkeit von der Beobachtungsdauer angegeben werden

$$\lambda_i(t) := r_i(t)\lambda(t)$$

Die individuelle Hazardrate wird über die Zeit unter Beobachtung hinaus 0.

Der individuelle kumulierte Intensitätsprozeß

$$\Lambda_i(t) = \int_0^t r_i(u)\lambda(u)du$$

gibt die erwartete Anzahl von Ausfällen an und nimmt nach Ende der individuellen Beobachtungsdauer nicht mehr zu.

Die Differenz zwischen beobachtetem Ausfallprozeß und kumuliertem Intensitätsprozeß zum Zeitpunkt t lautet

$$M_i(t) := \delta_i(t) - \Lambda_i(t)$$

und ist ebenfalls ab dem Ende der Beobachtungsdauer konstant, wenn ein absorbierender Zustand erreicht ist.

Martingal als binäre Verlustfunktion

Im Falle einer Vorhersage und ohne Zensierungen kann in der Notation der binären Verlustfunktionen $\tilde{S}_t(t_b) = \exp(-\tilde{\Lambda}_t(t_b))$ angegeben werden. Im Vergleich zu

$$\begin{aligned} L(t, \tilde{S}(t_b)) &= I_{\{t > t_b\}} - \tilde{S}_t(t_b) \\ &= I_{\{t > t_b\}} - \exp(-\tilde{\Lambda}_t(t_b)) \\ &= I_{\{t \leq t_b\}} - (1 - \exp(-\tilde{\Lambda}_t(t_b))) \\ &\neq I_{\{t \leq t_b\}} - (1 - \exp(-\tilde{\Lambda}_t(t_b))) \end{aligned}$$

kann dann $M_t(t_b)$ als Verlustfunktion geschrieben werden,

$$L(t, \tilde{S}_t(t_b)) = I_{\{t \leq t_b\}} - \tilde{\Lambda}_t(t_b)$$

Es wird dem beobachteten Zustand bei t_b statt der kumulierten Wahrscheinlichkeit für einen Ausfall bis t_b die kumulierte Anzahl von erwarteten Ausfällen bis t_b

gegenübergestellt. Dabei nimmt im Gegensatz zur Wahrscheinlichkeit die Anzahl der Ausfälle ab der abgelaufenen individuellen Beobachtungszeit t nicht mehr zu. Durch die Begrenzung bei der Vorhersage auf die individuelle Beobachtungsdauer innerhalb der Verlustfunktion ist eine separate Behandlung der Verluste von zensierten Elementen – wie bei der konservativen Verlustfunktion $L_{kons}((t_i, \delta_i), \tilde{t})$ – nicht notwendig. Eine Extrapolation für unvollständige Beobachtungen wird vermieden, da auch die Vorhersage entsprechend eingeschränkt wird. Jedoch ist die Vorhersage individuell, jedes Element erhält eine von seiner Beobachtungszeit abhängige Vorhersage der erwarteten Anzahl an Ausfällen.

Martingal-Residuum statt integrierter binärer Verlustfunktion

$M_i(t)$ ist ein Martingal dessen Summe beziehungsweise Mittelwert für jeden Zeitpunkt 0 ist. Mit dessen Hilfe kann über die gesamte Beobachtungsdauer integriert und so ein vielseitiges Residuum je nach Integrand erhalten werden. In seiner allgemeinen Form lautet es (Barlow&Prentice88)

$$\begin{aligned} e_i(f_i(t)) &= \int_0^{t_{max}} f_i(t) dM_i(t) \\ &= \int_0^{t_{max}} f_i(t) d\delta_i(t) - \int_0^{t_{max}} f_i(t) r_i(t) d\Lambda(t) \\ &= \int_0^{t_i} f_i(t) d\delta_i(t) - \int_0^{t_i} f_i(t) d\Lambda(t) \end{aligned}$$

Der einfachste Fall

$$\begin{aligned} e_i(1) &= \int_0^{t_i} d\delta_i(t) - \int_0^{t_i} d\Lambda(t) \\ &= \delta_i(t_i) - \Lambda_i(t_i) \\ &= \delta_i - \Lambda(t_i) \\ &= M_i(t_{max}) = M_i(t_i) =: M_i \end{aligned}$$

vergleicht die Anzahl der im Laufe seiner Beobachtungszeit aufgetretenen Ausfälle für Element i mit der Summe der im Laufe der Beobachtungszeit von Element i erwarteten Ausfälle.

Eine Analogie zur integrierten binären Verlustfunktion

$$\int_0^{t_{max}} L(t, \tilde{S}(t_b)) dt_b$$

besteht nicht mehr, da im nichtzensierten Fall $e_t(1)$ geschrieben werden kann als

$$\int_0^{t_{max}} dL(t, \tilde{S}_t(t_b)) = \int_0^t dL(t, \tilde{S}(t_b)) = L(t, \tilde{S}(t))$$

Es wird nicht der insgesamt erwartete Verlust, sondern der Verlust bei Ende der Beobachtung angegeben.

Das Martingal-Residuum M_i erreicht als Maximum 1 und als Minimum $-\infty$. Eine Möglichkeit, eine weniger schief verteilte Größe zu erhalten, ist eine Transformation zu Devianzresiduen (Therneau&al90)

$$dev_i := \text{sgn}(M_i)(-2(M_i + \delta_i \ln(\delta_i - M_i)))^{\frac{1}{2}}$$

Positive Residuen werden durch einen Logarithmus vergrößert und große negative Residuen durch eine Wurzeltransformation verkleinert. Bei semiparametrischer Modellannahme für das Cox-Modell mit $\lambda_i(t) = r_i(t) * \lambda_0(t) \exp(\beta x)$ und $M_i = \delta_i - \Lambda_0(t_i) \exp(\beta x)$ ergibt dev_i den Beitrag von Element i zur Devianz Dev_{PL} .

Allgemeines Martingal-Residuum zur Schätzung der EV

Ein Residuum, das ebenfalls nicht so schief verteilt ist wie $M_i = e_i(1)$ und den Verlust der Vorhersage nicht bezüglich Ausfall, sondern Ausfallzeit mißt, lautet

$$\begin{aligned} e_i(t) &= \int_0^{t_i} t d\delta_i(t) - \int_0^{t_i} t d\Lambda(t) \\ &= t_i \delta_i - \int_0^{t_i} t d\Lambda(t) \end{aligned}$$

Es vergleicht die tatsächliche Ausfallzeit eines Elements i mit der entsprechenden vorhergesagten Größe, unter der Bedingung, daß Element i bis t_i unter Risiko steht. Die beobachtete Ausfallzeit für ein zensiertes Element ist 0. Für Elemente mit kurzen Ausfallzeiten ergeben sich eher positive Residuen, Elemente mit später Ausfallzeit tendieren zu negativen Residuen (Abb.2).

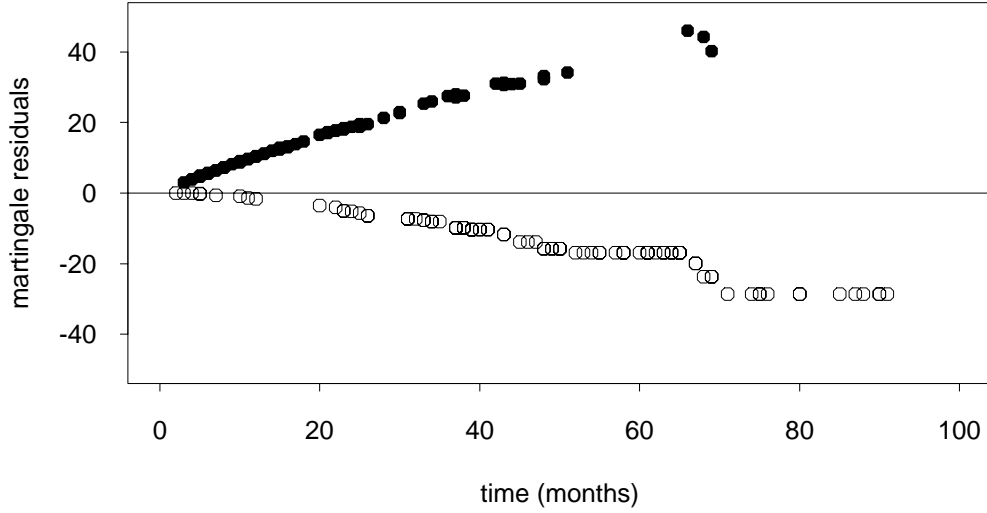


Abbildung 2: Martingal-Residuen $e_i(t)$ für Magenkarzinompatienten bei verteilungsfreier Schätzung $\bar{\Lambda}(\cdot)$. Kreise geben jeweils die Residuen für Zensierungen, Punkte für nichtzensierte Elemente an.

Es können unter Verwendung von

$$e_i^2(t) \equiv L((t_i, \delta_i), \tilde{t}_{0_{best}} | t_i)$$

beziehungsweise mit $\tilde{\Lambda}(t|x_i)$ für das Kovariablenmodell

$$e_i^2(t)|x_i \equiv L((t_i, \delta_i), \tilde{t}_{best} | t_i, x_i)$$

Verluste für das Nullmodell ebenso wie des Kovariablenmodell angegeben werden (Abb.3).

Ein Schätzer für den erwarteten Fehler im Modell mit beziehungsweise ohne Kovariablen ergibt sich als Mittel der beobachteten Fehler. Das Verhältnis der Fehler analog zu rv lautet

$$rv_{MR} := \frac{\frac{1}{n} \sum e_i^2(t) - \frac{1}{n} \sum e_i^2(t)|x_i}{\frac{1}{n} \sum e_i^2(t)}$$

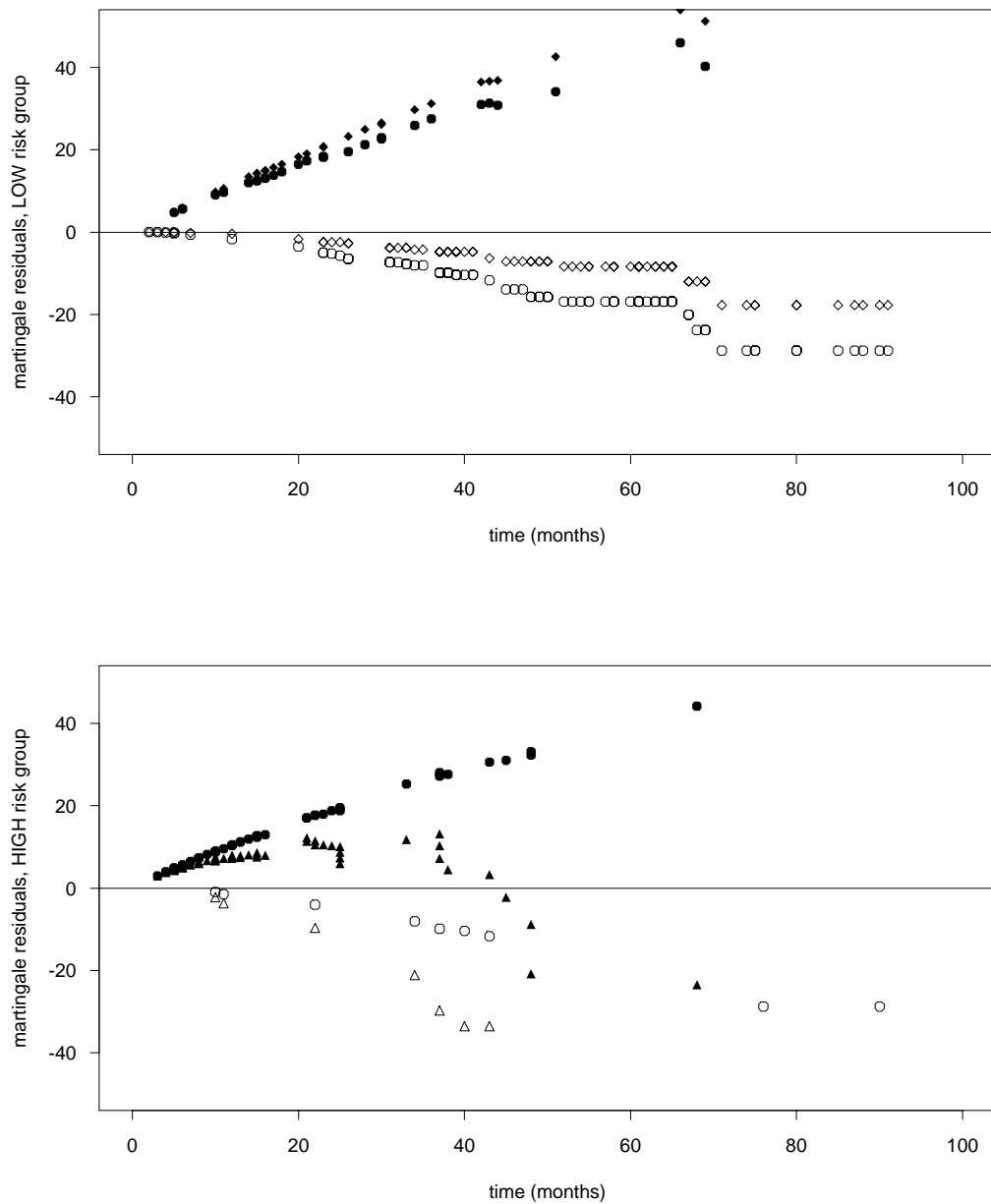


Abbildung 3: Martingale-Residuen für die Niedrig-Risikogruppe (oben, Symbol Raute) und Hoch-Risikogruppe (unten, Symbol Dreieck) bei Verwendung des verteilungsfrei geschätzten Kovariablenmodells. Im Vergleich Residuen derselben Patienten bei Verwendung des Nullmodells zur Vorhersage.

Somit läßt sich mit Hilfe der Martingalresiduen ein Schätzer für die EV bezüglich der vorhergesagten Überlebenszeit angeben, welcher nicht bei Zensierungen für deren Verluste extrapoliert, streicht oder geringer bewertet. Speziell für nichtparametrische Schätzungen ist so eine Beurteilung des Vorhersagewertes möglich, jedoch ist die Abhängigkeit von der angenommenen Zeit unter Beobachtung für die Vorhersage selbst problematisch.

7 Interpretation der geschätzten erklärten Variation

Wurde die Vorhersage parametrisch, das heißt unter Verwendung einer geschätzten Verteilung bestimmt, so kann die EV unter der Annahme berechnet werden, daß $\hat{F}(t|x)$ das korrekte Modell repräsentiert und $\hat{F}_0(t)$ das ebenfalls korrekte Null-Modell darstellt. \hat{EV} beziehungsweise \hat{EV}_{F_0} gibt dann einen Wert an, der für dieses Modell gilt, jedoch nicht unbedingt für die zugrunde liegenden Daten, da diese bei der Beurteilung nicht mehr verwendet werden. Ist die obige Annahme nicht erfüllt, das heißt die Anpassung des Modells an die Daten gering, ist \hat{EV} beziehungsweise \hat{EV}_{F_0} ein fiktiver Wert mit entsprechend geringem Bezug zu den Daten.

Der Fehler der Vorhersage im Nenner kann statt unter Verwendung von $\hat{F}_0(t)$ unter Verwendung von $\hat{F}(t)$ geschätzt werden. Für die Schätzung des Fehlers ohne Verwendung der Kovariablen wird dann ein Modell als korrekt angenommen, das nicht von der Existenz von Unterschieden in der Population ausgeht. Jedoch ist auch in der Anwendung bei Vernachlässigung der Kovariablen die Verwendung eines besser angepassten Modells wie $\hat{F}_0(t)$ die realistischere Alternative für einen Vergleich. Gleiches gilt für die Vorhersage selbst, jedoch ist \tilde{t}_{best} oft konsistenter Schätzer für \tilde{t}_{opt} .

Bei der Schätzung des Fehlers liefern die Verwendung von $\hat{F}(t)$ und $\hat{F}_0(t)$ unterschiedlich zu interpretierende Ergebnisse. Es wird nicht die durch die Kovariablen erklärte Variation gemessen, sondern die durch das Kovariablenmodell erklärte Variation, \hat{EV}_F . So kann die erklärte Variation von Kovariablenmodellen bei-

spielsweise bei unterschiedlichen Verteilungen verglichen werden.

Weiterhin kann sowohl bei Vorhersagen unter Verteilungsannahme als auch verteilungsfreien Vorhersagen die EV verteilungsfrei geschätzt werden. Bei Verwendung der empirischen Verteilungsfunktion und ohne das Vorliegen von zensierten Beobachtungen gilt wegen der optimalen Anpassung der empirischen Verteilung an die Daten anders als oben für die Verteilung im Nenner $\bar{F}_0(t) = \bar{F}(t)$. Also gilt für den empirischen Schätzer der erklärten Variation $ev_{\bar{F}_0} = ev_F = ev$. Ohne Zensierungen ergeben sich bei der verteilungsfreien Schätzung der erklärten Variation im Zähler und im Nenner die Residuen der Vorhersagen; die verteilungsfrei geschätzte EV ist die erklärte Residuen-Variation rv . Es werden die Fehler ohne Annahme eines wahren Modells geschätzt und so gleichzeitig die Anpassung der Daten an das gewählte Modell bestimmt. Deshalb kann sich bei nicht korrektem Modell, anders als oben, selbst bei $\hat{\beta} > 0$ im Extremfall $ev = 0$ ergeben. Diese Differenz zwischen \hat{EV} und ev kann bei der Schätzung des Erklärungsgrades der Kovariablen als Strafterm für die nicht optimale Anpassung des Modells interpretiert werden (Korn&Simon91). Ebenso kann ein Unterschied der unter Modellannahme und ohne Modellannahme geschätzten Fehler des Kovariablenmodells als ein Indikator für den Bedarf eines weiterentwickelten Regressionsmodells gelten (Graf&Schumacher95). Der analoge Vergleich der Fehler im Nenner gibt dann einen Hinweis auf die Anpassung nur bezüglich gewählter Verteilungsklasse und Response-Funktion. Ein gleicher Wert von ev und \hat{EV} weist auf eine korrekte und ausreichende Modellspezifikation hin und der Wert kann als relativer Gewinn bei der Präzision der Vorhersage aufgrund Trennschärfe und Verteilung der Kovariablen interpretiert werden.

Analog gibt ev allein keine Auskunft über die Güte der Anpassung des Modells, da sich selbst bei idealem Modell und perfekter Trennschärfe der Kovariablen aufgrund der ungünstigen Verteilung von X , beispielsweise diskreter Verteilung, eine niedrige Präzision der Vorhersage ergeben kann, die zu $ev < 1$ führt. Nur bei $\hat{EV} = 1$ wäre ev als Maß für die Güte des Modells zu interpretieren.

So können beste Kovariablenmodelle unterschiedlicher Modellklassen und Methoden bezüglich Erklärungsgrad der Kovariablen in diesem Modell, bezüglich des Kovariablenmodells insgesamt oder bezüglich der Anpassung des Modells vergli-

chen werden.

Literatur

BARLOW W.E., PRENTICE R.L. (1988). Residuals for relative risk regression. *Biometrika*, 75, 65-74.

COX D.R., SNELL E.J. (1989). *The Analysis of Binary Data*. Chapman & Hall, London.

FLEMING T.R., HARRINGTON D.P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

FRAZER D.A.S. (1965). On information in statistics. *Annals of Mathematical Statistics*, 36, 890-896.

GRAF E., SCHUMACHER M. (1995). An investigation on measures of explained variation in survival analysis. *The Statistician*, 44, 497-507.

HARRELL F.E. (1986). The PHGLM procedure. *SUGI Supplemental Library User's Guide*. Version 5 ed., SAS Institute Inc., Cary, NC, 437-448.

HARRELL F.E. (1997). *Predicting Outcomes: Applied Survival Analysis and Logistic Regression*. University of Virginia.

HENDERSON R. (1995). Problems and prediction in survival analysis. *Statistics in Medicine*, 14, 161-184.

KENT J.T. (1983). Information gain and a general measure of correlation. *Biometrika*, 70, 163-173.

KENT J.T. (1986). The underlying structure of nonnested hypothesis tests. *Biometrika*, 73, 333-343.

KENT J.T., O'QUIGLEY J. (1988). Measures of dependence for censored survival data. *Biometrika*, 75, 525-534.

KORN E.L., SIMON R. (1990). Measures of explained variation for survival data. *Statistics in Medicine*, 9, 487-503.

KORN E.L., SIMON R. (1991). Explained residual variation, explained risk and goodness of fit. *American Statistician*, 45, 201-206.

MADDALA G.S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge University Press.

MC CULLAGH P., NELDER J.A. (1983). *Generalized Linear Models*. Chapman & Hall, London.

MITTELBÖCK M., SCHEMPER M. (1996). Explained variation for logistic regression. *Statistics in Medicine*, 15, 1987-1997.

NAGELKERKE N.J.D. (1991). A note on the general definition of the coefficient of determination. *Biometrika*, 78, 691-692.

SCHEMPER M. (1990). The explained variation in proportional hazards regression. *Biometrika*, 77, 216-218.

SCHEMPER M. (1992). Further results on the explained variation in proportional hazards regression. *Biometrika*, 79, 202-204.

SCHEMPER M., STARE J. (1996). Explained Variation in survival analysis. *Statistics in Medicine*, 15, 1999-2012.

THEIL H. (1970). On the estimation of relationships involving qualitative variables. *American Journal of Sociology*, 76, 103-154.

THERNEAU T.M., GRAMBSCH P.M., FLEMING T.R. (1990). Martingale-based residuals for survival models. *Biometrika*, 77, 147-160.

VAN HOUWELINGEN J.C., LE CESSIE S. (1990). Predictive value of statistical models. *Statistics in Medicine*, 9, 1303-1325.